U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-256 679

Effect of Mechanical Cooling Devices on Ambient Salt Concentration

ADAPT Service Corp.

Prepared For
Pacific Northwest Environmental Research Lab.

April 1976

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/3-76-034	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Effect of Mechanical Cooling Devices on Ambient Salt Concentration		5. REPORT DATE April 1976 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Herbert E. Hunter		8. PERFORMING ORGANIZATION REPORT NO. ADAPT 75-8
9. PERFORMING ORGANIZATION NAME AND ADDRESS ADAPT Service Corporation 23 Pine Ridge Circle Reading, Massachusetts 01867		10. PROGRAM ELEMENT NO. 1BA032 11. CONTRACT/GRANT NO. 68-03-2176
12. SPONSORING AGENCY NAME AND ADDRESS EPA/Pacific Northwest Research Lab Corvallis, Oregon 97330		13. TYPE OF REPORT AND PERIOD COVERED Final-Feb 1975-Sept 1975 14. SPONSORING AGENCY CODE EPA-ORD

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report presents an analysis of the airborne salt concentration data collected during the demonstration of the salt water mechanical cooling devices at the Turkey Point power plant. The data were analyzed using the ADAPT family of empirical analysis programs which are based on the concept that empirical analysis should be preceded by the development of an optimal (in the Karhunen-Loeve sense) representation of the data. The analysis presented in the report shows that the increase in the background salt concentration due to the cooling tower was less than the measurement accuracy of approximately three to five micrograms per cubic meter. The analysis also shows that the spray modules used in this test probably increased the background concentration at one station located approximately 430 meters from the spray module by approximately three micrograms per cubic meter. These results were obtained by analysis of statistical summaries of the difference between the measured concentration with the cooling device operating and the calculated background concentration for the same conditions.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Airborne Salt Concentration	Cooling Towers	13/02
Regression Analysis	Spray Modules	06/06
Statistical Data Analysis		12/01
Thermal Pollution		04/02
		18/05
		20/04

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 142
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE 6.00

EPA Form 2220-1 (9-73)

PRICES SUBJECT TO CHANGE

☆ U.S. GOVERNMENT PRINTING OFFICE: 1976—697-053/83 REGION 10

[illegible]

EPA-600/3-76-034
April 1976

PB 256 679

Ecological Research Series

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ECOLOGICAL RESEARCH series. This series describes research on the effects of pollution on humans, plant and animal species, and materials. Problems are assessed for their long- and short-term influences. Investigations include formation, transport, and pathway studies to determine the fate of pollutants and their effects. This work provides the technical basis for setting standards to minimize undesirable changes in living organisms in the aquatic, terrestrial, and atmospheric environments.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/3-76-034
April 1976

EFFECT OF MECHANICAL COOLING DEVICES ON AMBIENT SALT CONCENTRATION

by
Herbert E. Hunter
ADAPT Service Corporation
Reading, Massachusetts

Contract 68-03-2176

Project Officer
Bruce A. Tichenor

Assessment and Criteria Development Division
Corvallis Environmental Research Laboratory
Corvallis, Oregon 97330

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
CORVALLIS ENVIRONMENTAL RESEARCH LABORATORY
CORVALLIS, OREGON 97330

ib

DISCLAIMER

This report has been reviewed by the Corvallis Environmental Research Laboratory, U. S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U. S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

CONTENTS

Page

List of Tables iv

List of Figures vi

I Introduction 1

II Conclusions and Recommendations 3

III Effect of Cooling Devices on Airborne Salt Concentration 8

Estimate of the Precision Run Error 8
Estimate of Background Concentrations 17
Effect of Cooling Tower on Ambient Concentration 38
Effect of Spray Modules on Ambient Concentration 40

IV Extension to Other Sites 47

V References 49

VI Appendices

Appendix A - Mathematical Description of ADAPT 50
B - Analysis of Optimal Bases Used for Salt Concentration Studies 99
C - Selection and Analysis of Algorithms 114
D - Algorithms for Calculating Ambient Concentration 124

LIST OF TABLES

Number Page

1 Definition of Data Vector 9

2 Most Important Environment Variables for Estimate of Precision Run Error 12

3 Order of Plot Symbols 16

4 Summary Concentration Statistics for Background Conditions 18

5 The Most Important Environmental Variables for Estimate of Background Salt Concentration (Based on East Wind Data Plus or Minus 45 Degrees) 18

6 The Most Important Environmental Variables for Estimate of Background Salt Concentration (Based on Absolute East Wind Data) 25

7 The Most Importance Environmental Variables for Estimate of Background Concentration at Stations 1 and 2 25

8 The Most Important Environmental Variables for Estimate of Background Concentration at Station 3 30

9 The Most Important Environmental Variables for Estimate of Background Concentration at Station 4 30

10 The Most Important Environmental Variables for Estimate of Background Concentration at Station 5 31

11 The Most Important Environmental Variables for Estimate of Background Concentration at Station 6 31

12 The Most Important Environmental Variables for Estimate of Background Concentration at Station 7 32

13 The Most Important Environmental Variables for Estimate of Background Concentration at Station 8 32

14 The Most Important Environmental Variables for Estimate of Background Concentration at Station 9 33

15 The Most Important Environmental Variables for Estimate of Background Concentration at Station 10 33

LIST OF TABLES (CONT'D)

Number		Page
16	The Most Important Environmental Variables for Estimate of Background Concentration at Station 11	34
17	Summary Statistics for Down Wind Minus Background Concentration for Cooling Tower	39
18	Summary Statistics for Down Wind Minus Background Concentration for Spray Modules	39
19	The Most Important Environmental Variables for Estimate of Increase in Salt Concentration Due to Spray Modules at Station 7	45
20	The Most Important Environmental Variables for Estimate of Increase in Salt Concentration Due to Spray Modules at Stations 6 through 9	45

LIST OF FIGURES

Number		Page
1	Airborne Particles Sampler Station Locations	4
2	Relative Importance of Indexing Variable Defined by Table 1 to Estimate of Precision Run Error Using an Eleven Dimensional Analysis	11
3	Estimated Versus Actual Precision Run Error	14
4	Estimated Versus Actual Ambient Salt Concentration Pooled Over All Measurement Stations	19
5	Definition of Wind Vector	20
6	Relative Importance of Indexing Variable Defined by Table 1 to Estimate of Ambient Salt Concentration Pooled Over All Stations	22
7	Relative Importance Vector for East Wind (-45°)	25
8	Relative Importance Vector for Absolute East Wind (-45°)	26
9	Relative Importance Vector for Ambient Concentration at Station 10	35
10	Relative Importance Vector for Ambient Concentration at Station 9	36
11	Relative Importance Vector for Spray Modules at Station 7	43
12	Relative Importance Vector for Spray Modules at Stations 6 through 9	44

SECTION I

INTRODUCTION

The purpose of this study is to analyze the airborne salt concentration data collected during the demonstration of salt water mechanical cooling devices at the Turkey Point power plant and reported in detail in Reference 1. Airborne particle samplers were used to collect data on ambient salt concentration. The purpose of the study reported on Reference 1 was to measure the amount of cooling device draft which was emitted from the cooling devices and subsequently collected at downward samplers. The data consist of a series of measurements to define the background airborne salt concentration at each of the stations shown in Figure 1. Note, that Stations 1 and 2 are co-located and thus provide data on the repeatability of the measurements or the "precision run error". A second set of data were collected at Stations 3 through 11 during the operation of either a single cell salt water cooling tower or a pair of Powered Spray Modules. For all of the data runs, only one of the two types of cooling devices was operating so that the data may be divided into three classes: 1) background salt concentration, 2) cooling tower plus background salt concentration and 3) powered spray modules plus background salt concentration. By definition, background data define the conditions with no cooling device in operation.

In general, the early portion of the data collection consisted of collecting background data only. This situation existed from approximately August of 1973 through early January of 1974. During this time period, only measurement Stations 1 through 6 were used. Beginning at the end of January 1974, measurements were made using each of the cooling devices independently as well as a few background measurements. During this time period, Stations 3 through 11 were used. The last series of background measurements were made in April of 1974. Between April of 1974 and the end of July of 1974, all of the measurements were made with one of the cooling devices operating.

The approach to the analysis of these data consisted of utilizing the data obtained at the two co-located stations to make an estimate of the precision of the measurements. The background data were then used to develop regression algorithms relating the background salt concentration to the environmental variables. These algorithms were then used to estimate the background concentrations at the time that the data were being collected during the operation of the cooling devices. This estimated background concentration was then subtracted from the measured concentration during the operation of the cooling device to determine the effect of the cooling device on the ambient concentration. Regression analysis was then performed to relate this effect of the cooling device on the ambient concentration to the environmental variables.

The ADAPT family of computer programs, based on the concept of preceding empirical analysis with the transformation of the data to the principle component space offers a unique capability to analyze these data. This transformation not only reduces the amount of computation required to analyze the data, but also provides analysis of the data which is useful for understanding the structure of the data. This analysis will insure that no major errors have been made in the preparation of the data. The advantages as well as a detailed description of the ADAPT approach for performing the regression analyses required for this study are summarized in Appendix A.

SECTION II

CONCLUSIONS AND RECOMMENDATIONS

This section will present a brief summary of the results and recommendations resulting from the studies of the Turkey Point data. Justification for these results and recommendations are presented in Section 3 and the supporting appendicies.

EFFECT OF COOLING DEVICES ON BACKGROUND SALT CONCENTRATION

The primary results of the present study are: 1) the operation of the cooling tower at the Turkey Point power plant did not increase the background salt concentration by a measurable amount at any of the stations, and 2) the effect of operating the spray modules at the Turkey Point power station probably increased the background salt concentration at Station 7 by approximately 50% and did not increase it by a measurable amount at any other station.

These results are developed in detail in Sections 3.3 and 3.4. They are based on the results of statistical summaries of the difference between the concentration measured with the device operating and the expected background concentration made at each station and pooled over all stations. The average difference between the measured concentration with the cooling tower operating and the expected background concentration was the order of 2×10^{-3} micrograms per cubic meter with a standard deviation of 4.8 micrograms per cubic meter. Similar results obtained for each of the individual stations indicate there were no stations for which the difference between the measured concentration with the cooling tower operating and the expected background concentration averaged over all of the measurements at the station exceeded the standard deviation.

The average of all of the concentrations obtained during the operation of the spray modules, minus the expected background concentration is 1.32 micrograms per cubic meter with a standard deviation of 9.9 micrograms per cubic meter. When considering the individual stations, the average difference between the concentration measured with the spray modules operating and the expected background concentration only exceeded the standard deviation for two stations. One of these stations was Station 10 for which only 6 cases were available and thus no meaningful conclusions can be based on the results obtained at this station. The other station was Station 7 which Figure 1

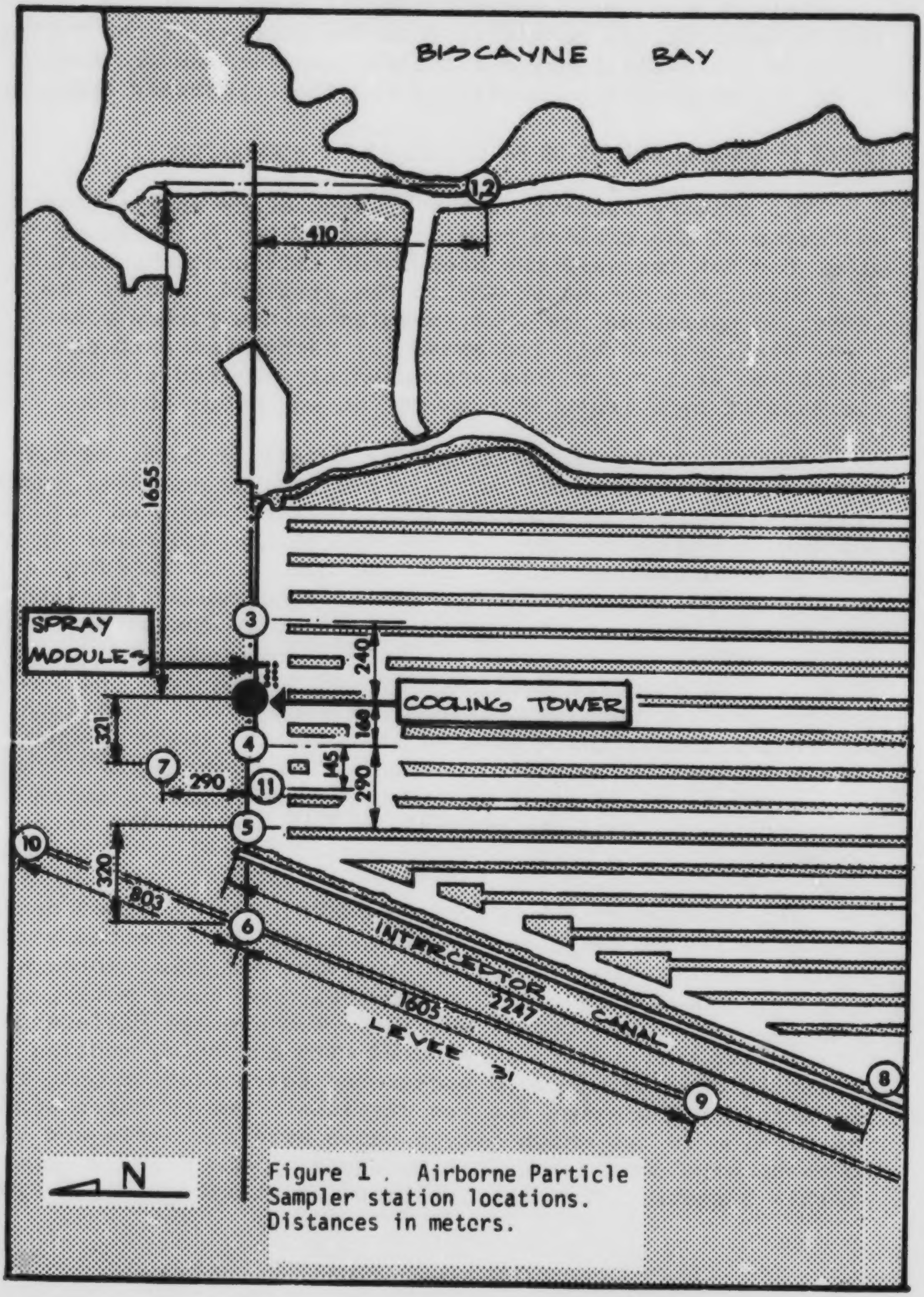

BISCAYNE BAY
12
410
1655
SPRAY MODULES
3
240
COOLING TOWER
160
321
4
7
145
290
11
290
5
10
320
803
6
INTERCEPTOR CANAL
1603
2247
LEVEE 31
9
8
N

Figure 1 . Airborne Particle Sampler station locations. Distances in meters.

shows to be one of the three closest stations to the spray modules. The standard deviation of the measurements at this station was smaller than any of the other nearby stations indicating that the measurement errors at this station were unusually small. For this special situation of a nearby station with an unusually small measurement error, there is approximately 85% confidence that the spray module has increased the expected background salt concentration. For this station, the most likely effect is that on the average the spray module increases the background salt concentration from approximately 5 to 8 micrograms per cubic meter.

It was not possible to prepare algorithms for estimating the effect of the cooling devices on the salt concentration as a function of environment and position because of the small effects of the cooling devices on the salt concentration. A more specialized optimal base may allow development of an algorithm for estimating the effect of the environment on the salt concentration at a fixed position; namely, that defined by Station 7. Algorithms were developed to calculate the background concentration at each station as a function of the environment. These algorithms are given in Appendix D. Thus, the major usefulness of the present study is: 1) the general result that the effect of these cooling devices is small compared to the measurement accuracy and 2) the information which can be gained by examining the ADAPT analysis outputs to determine the relative importance of the environmental variables to the concentrations measured and to define requirements for additional measurements.

RECOMMENDATIONS

The results of this study which are presented in Sections 3 and 4 of this report have led to the following specific recommendations:

1. For future tests in which it is desired to determine the precision run error, data should be obtained on co-located stations for at least 300 days. This amount of data is required in order to provide sufficient number of cases to develop an algorithm to determine the precision run error as a function of the environment.

2. Co-located stations should be provided at a number of different locations in order to allow the development of algorithms to determine the effect of location on the precision run error.

3. Additional emphasis should be placed on checking the equipment and motivating test personnel on Mondays or after any period of in-activity.

4. For those analyses where the major objective of the study is to determine the effect of the environment on the background salt concentration, the wind vector should be defined as the projection of the wind on a compass direction rather than on the position direction. Thus, separate ADAPT bases will be required for developing algorithms to study the effect of the environment and location on the background concentration and to determine the effect of cooling devices on this concentration as a function of location and environment.

5. At least 250 to 300 measurements are required at each station to adequately define the effect of environment on the background salt concentration.

6. In planning future tests programs the location of the stations relative to open water, including cooling basins, should be considered with a significant number of the stations located at distances greater than 300 meters from such open water.

7. A new ADAPT optimal base should be developed using only the data from Stations 6 through 9 and a second optimal base using only the data from Station 7. These bases should then be used to rederive the algorithms for estimating the increase in background concentration due to the spray modules at these stations.

SECTION III

EFFECT OF COOLING DEVICES ON SALT CONCENTRATION

The use of the Turkey Point data to estimate the effect of two cooling devices on the ambient, airborne salt concentration may be divided into the following three steps: 1) the estimation of the precision of the measurements, 2) an estimate of the ambient concentration had the device not been operating i.e. background concentration and 3) the statistical analysis of the difference between the actual measured concentration with the cooling device operating and the expected background concentration. The first two of these steps are identical for both the cooling tower and the spray modules operating and the final step was carried out independently for the data obtained when the cooling tower was operating and the data obtained when the spray modules were operating.

ESTIMATION OF THE PRECISION RUN ERROR

The precision run error is estimated from the results obtained when no cooling device was operating and measurements were made at Stations 1 and 2 which Figure 1 shows were co-located. These measurements were made for 65 different cases. The average precision run error for these 65 cases was 6.11% with the standard deviation of 9.55%. Thus, if the distribution is Gaussian we have a 70% confidence that the precision run error was between 2% and 10%.

A regression analysis was performed to develop an algorithm for estimating the precision run error as a function of the environmental variables. The independent variables which were used to prepare the estimate of the precision run error are defined in column three, headed PRE, of Table 1. The importance of each of these variables to the estimate of the precision run error is shown in the relative importance plot presented in Figure 2. The ordinate of Figure 2 is the relative importance of each of the environmental variables to the estimate of the precision run error. The number of learning cases was insufficient and covered an insufficient variation in environmental conditions to allow one to develop an algorithm having sufficiently high confidence and accuracy to warrant the application to the test conditions. However, the data were adequate to provide an indication of which environmental factors have the greatest impact on the precision run error. Figure 2 presents this information. The absolute magnitude of the

TABLE 1
DEFINITION OF DATA VECTOR

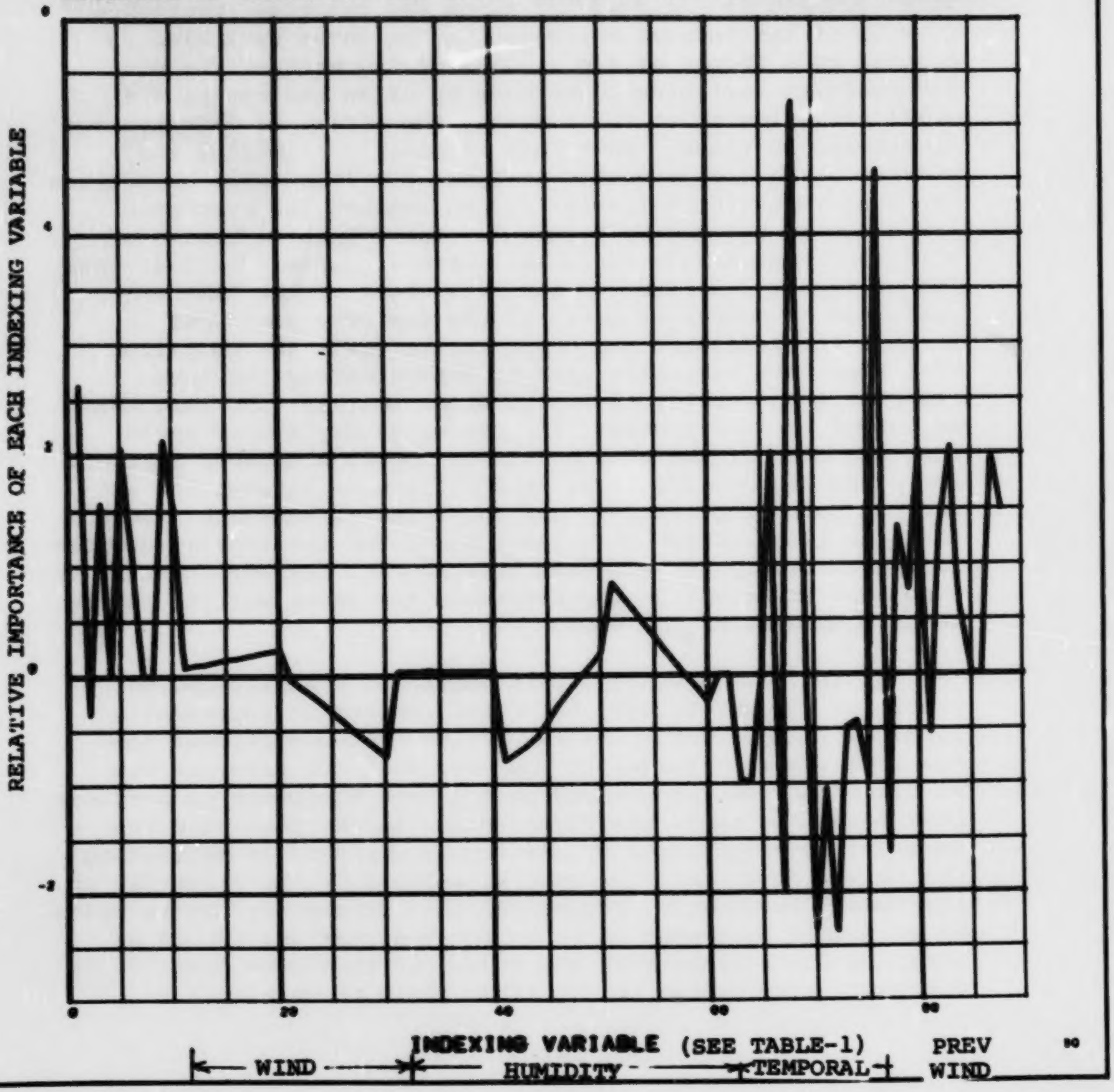
VARIABLE NO			SYMBOL	DESCRIPTION
82Pt	75Pt	PRE		
1	-	1(80)	C	Mass of Salt/Unit Volume of Air
2	-	2(81)	VOL	Sampled Air Volume, M^3
3	-	3(82)	N_a	Mass of Sodium on Mesh Pair
4	1	-	d_e	Projection of position vector on East direction
5	2	-	d_N	Projection of position vector on North direction
6	3	4	CC1	Binary Code for Light Rain
7	4	5(83)	CC3	Binary Code for Bugs on Sample
8	5	6(84)	CC4	Binary Code for Dust Contamination
9	6	7(85)	CC5	Binary Code for Combination of Comments
10	7	8(86)	CC9	Binary Code for White Caps
11	8	9(87)	ts	Start Time
12	9	10(88)	te	End Time
13-22	10-19	11-22	dwi (i=1,10)	Projection of Wind Vector on Position Direction -10 Samples Between ts and te
23-32	20-29	21-30	Nwi (i=1,10)	Projection of Wind Vector on Normal to Position Direction -10 Samples Between ts and te
33-42	30-39	31-40	Ti (i=1,10)	Dry Bulb Temperature -10 Samples Between ts and te
43-52	40-49	41-50	Di (i=1,10)	Difference Between Dry and Wet Bulb Temperature -10 Samples Between ts and te
53-62	50-59	51-60	Hi (i=1,10)	Relative Humidity -10 Samples Between ts and te
63	-	61	CDC1	Binary Variable Indicating Cooling Tower Operation
64	-	62	CDC2	Binary Variable Indicating Spray Modules Operating
65	60	63	DY	Day of Year
66	61	64	DFT	Days Since First Test
67	62	65	S1	Binary Variable Indicating Spring
68	63	66	S2	Binary Variable Indicating Summer
69	64	67	S3	Binary Variable Indicating Fall
70	65	68	M	Binary Variable-Test on Monday
71	66	69	T_u	Binary Variable-Test on Tuesday
72	67	70	W	Binary Variable-Test on Wednesday

( ) - Variable Taken from STA-2 for Pre Estimate

Table 1, continued

VARIABLE NO			SYMBOL	DESCRIPTION
82Pt	75Pt	PRE		
73	68	71	Th	Binary Variable-Test on Thursday
74	69	72	F	Binary Variable-Test on Friday
75	70	73	S	Binary Variable-Test on Saturday
76	71	74	$dw_{(-1)}$	Projection of Preceding Day's Average Wind Vector on Position Direction
77	72	75	$dn_{(-1)}$	Projection of Preceding Day's Average Wind Vector on Normal to Position Direction
78	73	76	$dwsp_{(-1)}$	Preceding Day's Spread in Wind Speed
79	74	77	$dnsp_{(-1)}$	Preceding Day's Spread in Wind Direction
80	75	78	$dw_{(-1)}$	Preceding Day's Standard Deviation of Wind Speed
81	-	79	$du_{(-1)}$	Preceding Day's Standard Deviation of Wind Direction
82	-	-	PRE	Precision Run Error

FIGURE 2 - RELATIVE IMPORTANCE OF INDEXING VARIABLE DEFINED BY TABLE 1 TO ESTIMATE OF PRECISION RUN ERROR USING AN 11 DIMENSIONAL ANALYSIS



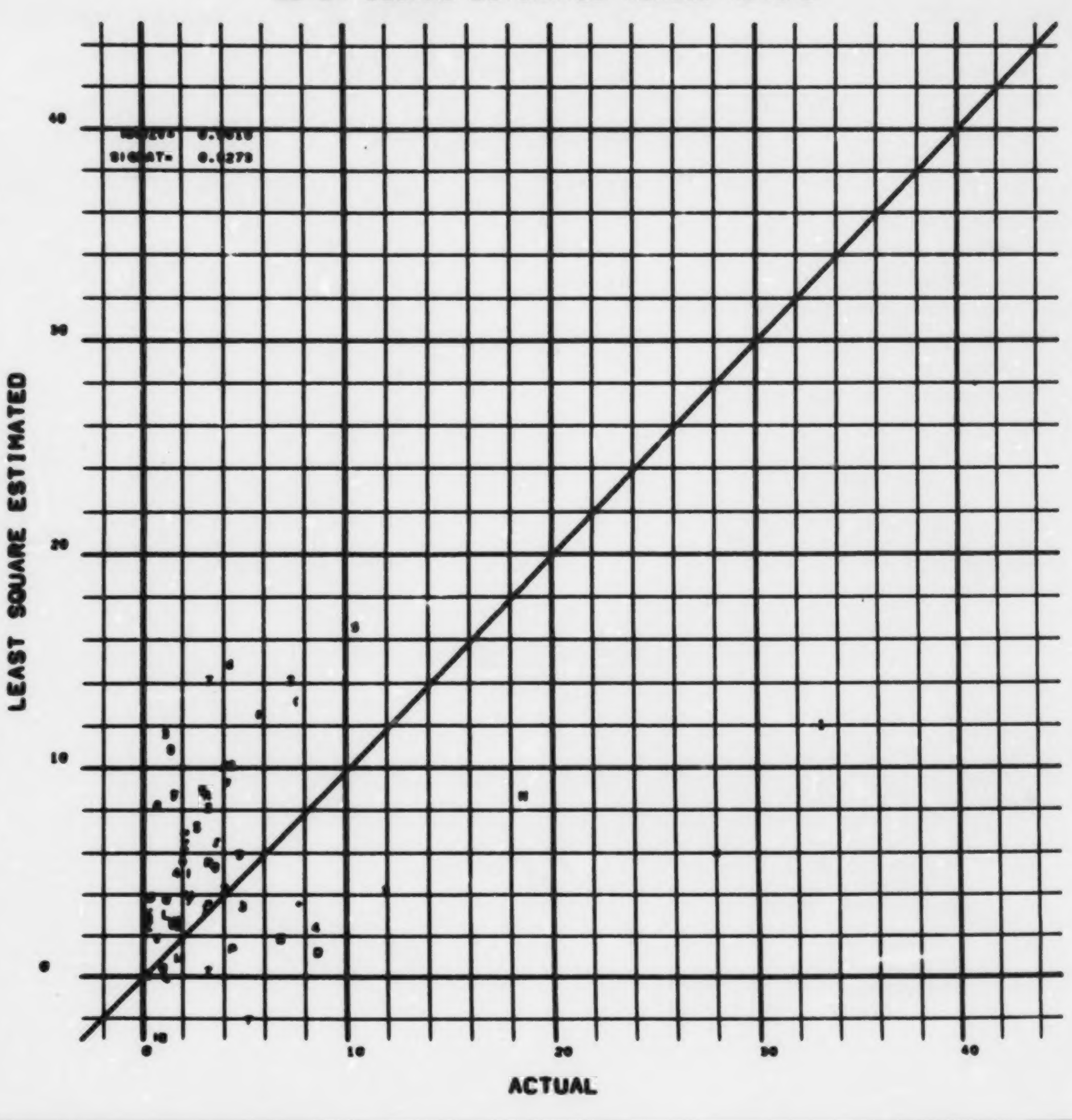
ordinate indicates the effect of each of these variables. One must also account for correlation between variables in interpreting this figure. For these data, this may be accomplished by multiplying the average value of each of the sets of variables representing the wind and the humidity (i.e., variables 11 through 60 taken in sets of 10) by 10. The need for this procedure can be seen by noting that if the same variable is included "n" times the relative importance of each of the indices corresponding to these variables is reduced by a factor of 1/n . Considering each of the wind and humidity variables to be made up of an average plus a small variation about the average, the effect of this average is entered 10 times. When this is done, one obtains the results which are presented in Table 2. This table summarizes the most important variables for estimating the precision run error as defined by Figure 3. Table 2 shows that the most important single variable (which accounted for less than 10% of the explained variation) was wheather the test was performed on Monday or not. If the test was performed on Monday, the precision run error was larger. The next four most important variables each of approximately the same importance and nearly as important as whether the test was performed on Monday were: the preceding day's wind speed, the component of the wind normal to the line between Stations 1 and 2 and the cooling tower, the difference between the wet and dry bulb temperatures, and the relative humidity. Note, that the line between Stations 1 and 2 and the cooling devices is approximately an east-west line so that the component of the wind most important to the precision run error was the north-south component of the wind.

Of the four most important variables contributing to precision run error, only the first appears to represent a factor which can easily be controlled. If one assumes that the importance of Monday to the precision run error is due to the effect of the weekend on the test equipment or personnel performing the test, some improvement in the precision run error could be achieved by particular emphasis on motivating the personnel to exercise special efforts on the first day of the week. It is also recommended that should any future tests be performed, information on co-located stations should be obtained for a minimum of 300 days. If this were done, it is extremely likely that an algorithm could be developed to

TABLE 2 - MOST IMPORTANCE ENVIRONMENT VARIABLES FOR ESTIMATE OF PRECISION RUN ERROR

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE INCREASE	DECREASE
Test on Monday	68	5	X	
Preceding Days Spread in Wind Speed	76	4½	X	
Wind Normal to Position Direction	21-30	4		X
Difference in Dry & Wet Bulb Temperature	41-50	4		X
Relative Humidity	51-60	4	X	
Salt Concentration Deposited	1	2½	X	
Bugs on Sample	5	2	X	
Start Time	9	2	X	
Test Performed in the Fall	67	2		X
Test on Wednesday	70	2		X
Test on Friday	72	2		X

FIGURE 3 - ESTIMATED VERSUS ACTUAL PRECISION RUN ERROR



estimate the precision run error as a function of the environmental variables shown in Table 1.

The results of applying the algorithm represented by the relative importance vector in Figure 2 to the data used to derive the algorithm is shown in Figure 3. Note, that this is a plot of the data which was used to derive the algorithm rather than independent test data. These data are often referred to as learning or training data. Thus, Figure 3 presents a plot of the estimated precision run error versus the actual precision run error for the learning data. The ordinate of this figure is the precision run error estimated using the regression algorithm corresponding to the relative importance vector of Figure 2. The abcissa of Figure 3 is the actual precision run error for each of these cases. The performance shown on this figure corresponds to a correlation coefficient of 0.56 which is equivalent to explaining approximately 17% of the variation in the data.

Each of the points shown on Figure 3 may be related to the raw data presented in Appendix C of Reference 1 by use of Table 3 which indicates the sequential order in which the symbols appear on the plot. The cases were plotted in chronological order according to this sequence. Since only 17% of the precision run error could be explained by the environmental factors, no attempt was made to estimate precision run error for the other environmental conditions occurring during the later tests. All of the learning data were obtained during the early portions of the test. The ADAPT analysis of the data which is reported in Appendix 2 showed that the character of the environment for the data obtained during the early portions of the experiment in which the precision run error was obtained was significantly different from the character of the environment during the later portions of the experiment. Since only 17% of the variation could be explained by the environmental factors, the best approach is to use 6% as the estimate of the precision run error with a 70% confidence that the actual precision run error lies between 2% and 10%. The analysis of the precision run error regression algorithm discussed in Appendix C of this report indicates that the most likely explanation of the inability to develop a successful algorithm for estimating the precision run error was the fact that only 65 learning cases were available for this estimate.

TABLE 3

ORDER OF PLOT SYMBOLS. READ FROM LEFT TO RIGHT

123456789=+-*/.()♦∂"δα?βπ•γ~◢∫

Σ□ABCDEFGHIJKLMNOPQRSTUVWXYZ

ESTIMATION OF BACKGROUND CONCENTRATIONS

The Turkey Point data were processed through the ADAPT programs to develop regression algorithms to estimate the background concentrations, (i.e., concentrations with no cooling device operating), as a function of the environmental variables defined in Table 1 for each of the measurement stations. These estimates were made for all of the stations pooled together, for all of the stations during east winds, and for each of the individual stations. The results of these estimates are summarized in Table 4. Physical reasoning indicates that the cooling device will only effect downwind stations, therefore, the background cases consisted of both the data obtained when the cooling devices were not operating and the upwind data.

Table 4 provides a summary of both the number of learning cases and the number of dimensions used for each algorithm, the performance of each algorithm in terms of the correlation coefficient and the explained variation and the mean concentrations and standard deviations of the learning data used to derive each of the algorithms. Each of the three sets of algorithms developed will be discussed independently in the following sections.

Estimate of Background Salt Concentration Pooled Over All Stations

The algorithm derived to estimate the background salt concentration using the data pooled over all of the stations proved to have a relatively poor performance as indicated by the correlation coefficient of 0.54 which provides an explained variation of 16%. Figure 4 shows the performance of this algorithm on each of the 478 learning cases used to derive the algorithm. This figure shows that those cases having very large salt concentrations were badly underestimated using this algorithm. It was hypothesized that the reason for this poor performance is due to the format of the wind vector.

The wind vector used was selected to be optimal for estimating the concentration due to the cooling devices and not the background concentration. To avoid the discontinuous change in direction occurring as one moves from 360 to 0 degrees occurring in a polar coordinate system, the wind vector used was the projection on two perpendicular directions as shown in Figure 5. Since the primary objective of this study is to

TABLE 4 - SUMMARY CONCENTRATION STATISTICS FOR BACKGROUND CONDITIONS

STATION	NO OF LEARN CASES	NO OF DIM USED	CORR COEF	EXPL VAR.	MEAN LEARN	CONC. DWN. WD (ug /m^3)	STD DEV. LEARN	DWN WIND
ALL-POOLED	478	30	0.54	0.16	6.41	-	4.80	-
POOLED EAST WD	181	20	0.65	0.25	6.29	-	4.42	-
POOLED ABSOLUTE EAST WIND	181	20	0.78	0.38	6.29	-	4.42	-
1,2	140	20	0.74	0.33	6.36	-	4.60	-
3	181	16	0.62	0.22	6.12	6.4	5.18	2.5
4	136	16	0.53	0.15	5.80	4.9	5.33	2.5
5	135	16	0.57	0.18	4.91	5.8	4.15	2.1
6	137	16	0.69	0.28	5.46	7.8	4.38	2.9
7	122	12	0.63	0.23	5.04	1.8	4.33	2.6
8	31	4	0.64	0.23	5.66	7.1	5.21	2.7
9	95	8	0.63	0.23	4.14	5.39	3.48	2.4
10	127	12	0.77	0.36	3.92	6.4	3.42	1.7
11	46	4	0.26	0.03	7.39	4.4	8.68	2.1

TABLE 5 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND SALT CONCENTRATION (BASED ON EAST WIND DATA PLUS OR MINUS 45°)

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRECISION INCREASE	DECREASE
Preceding Day's Spread and Standard Deviation of Wind Speed	73,75	46		X
Projection of Wind Vector on Position Direction	10-19	15	X	
Dry Bulb Temperature	30-39	10		X
Presence of White Caps	7	9	X	
Humidity	50-59	8		X
Projection of Position on East Direction	1	6	X	

FIGURE 4 - ESTIMATED VERSUS ACTUAL AMBIENT SALT CONCENTRATION POOLED OVER ALL MEASUREMENT STATIONS

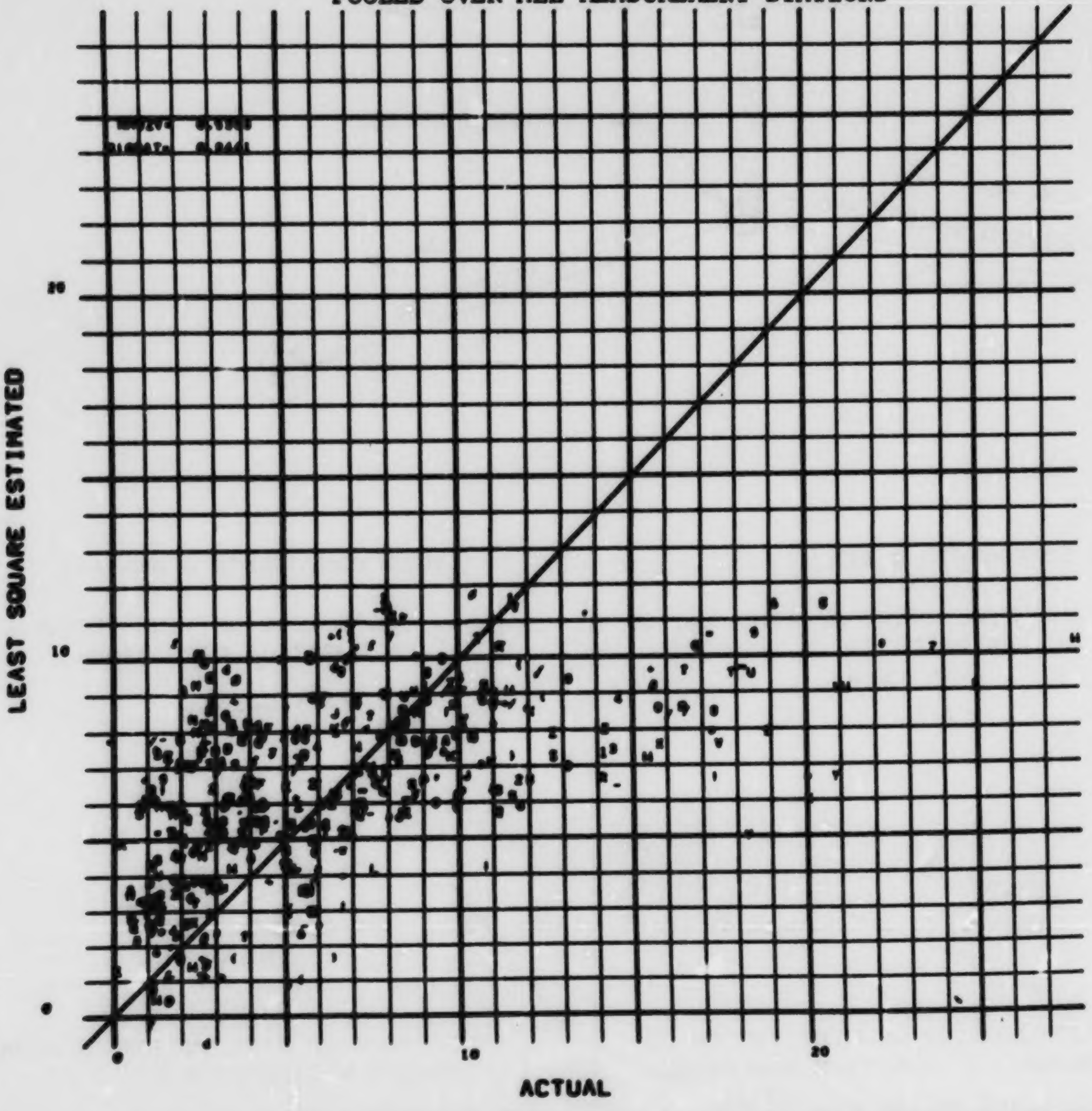
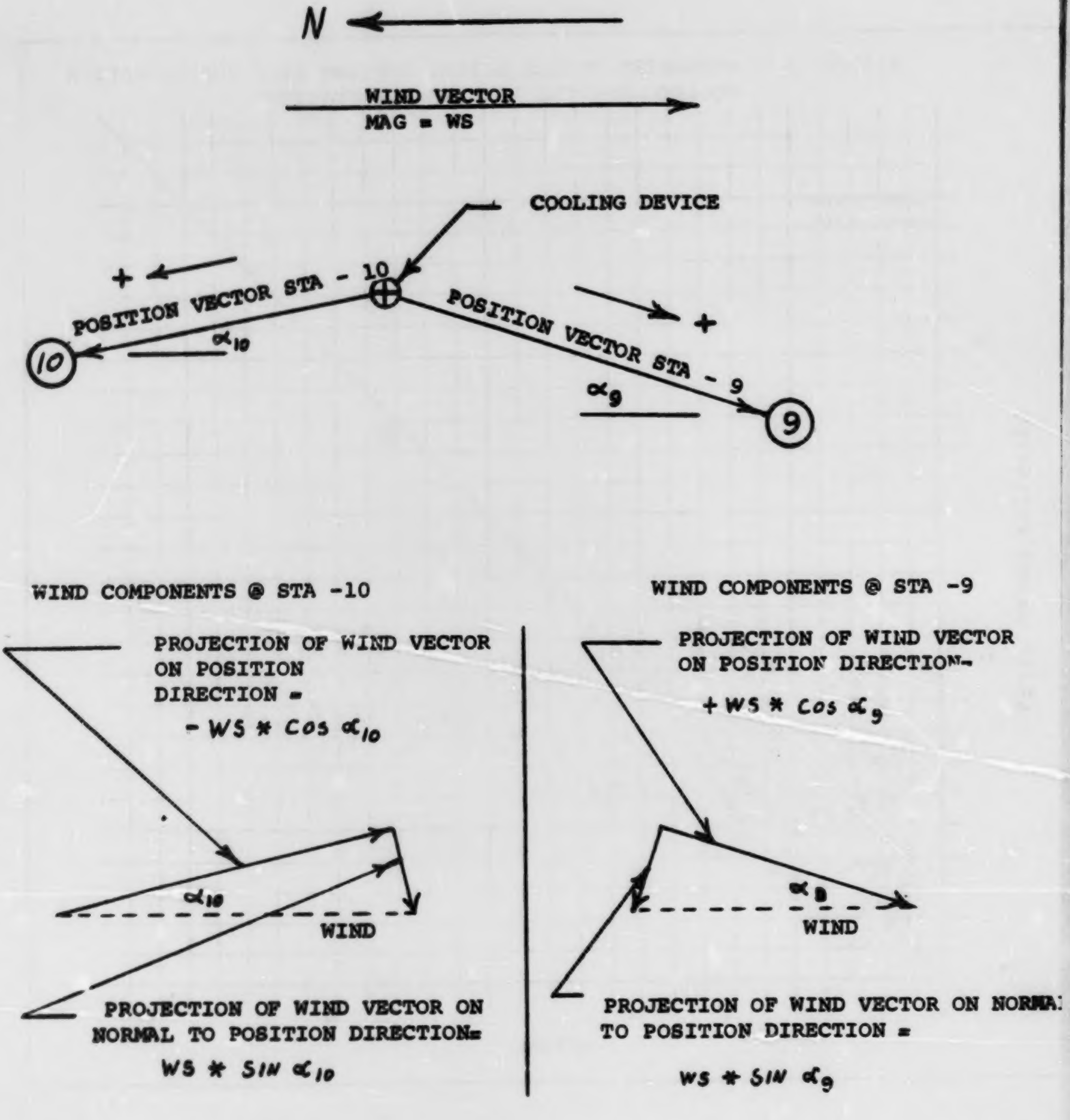


FIGURE 5 - DEFINITION OF WIND VECTOR
(ILLUSTRATION FOR NORTH WIND)

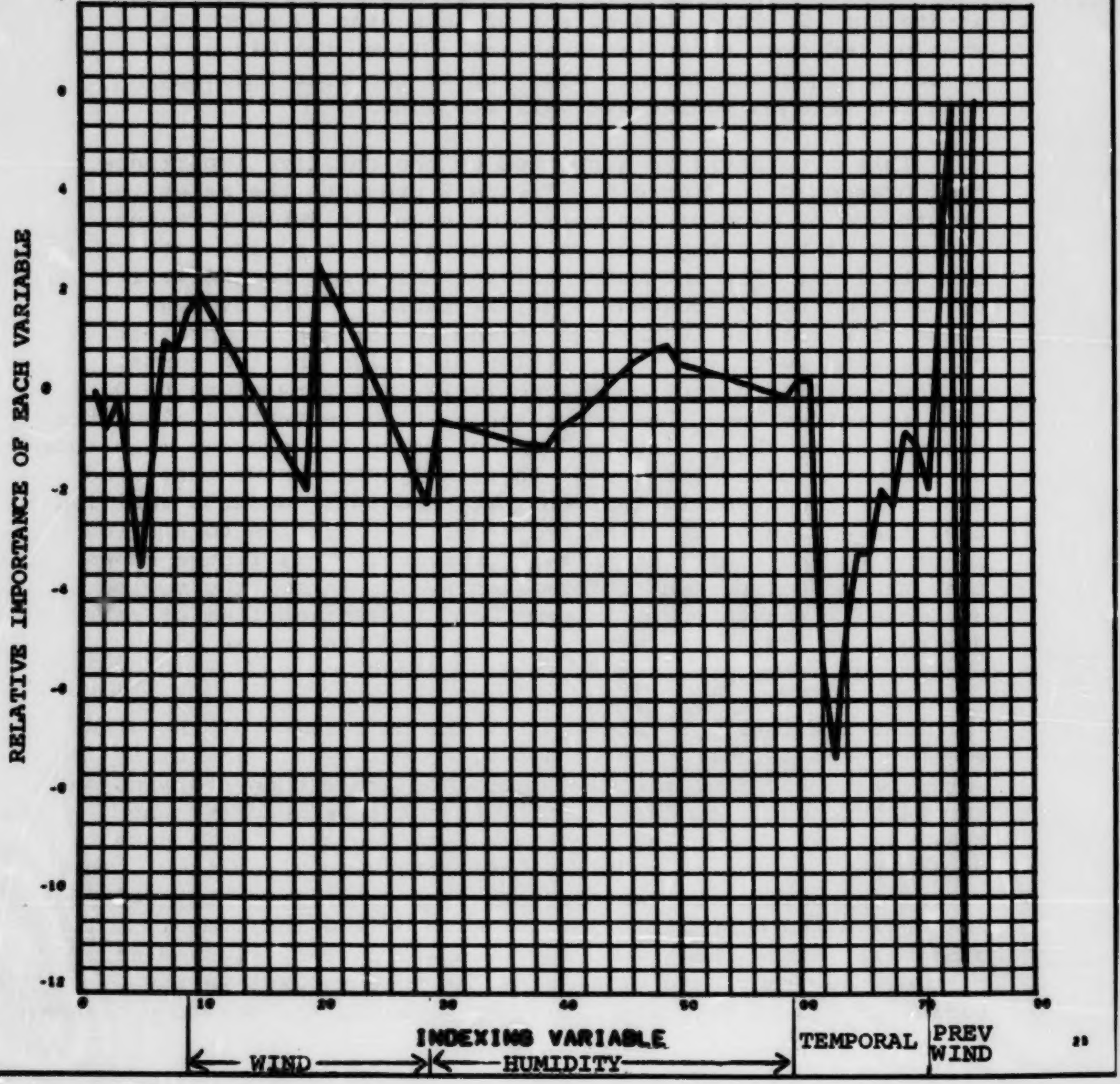

N
WIND VECTOR
MAG = WS
COOLING DEVICE
+
POSITION VECTOR STA - 10
α_{10}
10
POSITION VECTOR STA - 9
+
α_9
9
WIND COMPONENTS @ STA -10
PROJECTION OF WIND VECTOR ON POSITION DIRECTION =
$-WS * \cos\alpha_{10}$
α_{10}
WIND
PROJECTION OF WIND VECTOR ON NORMAL TO POSITION DIRECTION=
$WS * \sin\alpha_{10}$
WIND COMPONENTS @ STA -9
PROJECTION OF WIND VECTOR ON POSITION DIRECTION-
$+WS * \cos\alpha_9$
α_9
WIND
PROJECTION OF WIND VECTOR ON NORMA
TO POSITION DIRECTION =
$WS * \sin\alpha_9$

determine the effect of the cooling devices, the wind is projected on a coordinate system located relative to the line connecting the cooling devices and the measurement stations. Since measurement stations were located on all sides of the cooling devices, this coordinate system only has meaning with respect to the background salt concentration on a station by station basis. For example, Figure 5 shows that a North wind will be positive at Station 9 but negative at Station 10. Thus, the effect of this wind is cancelled when pooled over these two stations. Thus, the linear regression algorithm can not make use of the wind information when the data are pooled over all stations.

This hypothesis is supported by the relative importance plot presented in Figure 6. Figure 6 presents the relative importance of each of the variables shown in Table 1 to the estimate of the background salt concentration pooled over all stations. The environmental variables may be identified by using Column 2, headed 75 pt, of Table 1. The average value of variables 10 through 19 and 20 through 29, projection of the wind vector on the position direction and the normal of the position direction respectively, are very close to 0. This indicates that the background salt concentration algorithm derived using the pooled data is not using the average magnitude of the wind during the test. Physically, one knows that wind should be important to this estimate and thus, we have verified that the change of coordinate systems from station to station is denying the algorithm this information. This figure shows that the most important information available for the pooled estimate is the temporal information, indicating time of year during which the tests were performed and the spread of variation in the previous day's wind speeds.

The difficulty resulting from the definition of the wind vector could be corrected in two different ways. The first would be to redefine the wind vector by projecting the wind on two perpendicular compass directions and then rederive the ADAPT optimal base using these new data vectors. This base could then be used to rederive the regression algorithms. The second approach is to use the data vectors as presently defined and derive the concentrations for each of the individual stations. Since the second approach was already planned as part of the study, and eliminates the need to explain the variation due to position, this approach was the solution chosen.

FIGURE 6 - RELATIVE IMPORTANCE OF INDEXING VARIABLE DEFINED BY TABLE I TO ESTIMATE OF AMBIENT SALT CONCENTRATION POOLED OVER ALL STATIONS



Estimate of Background Salt Concentration Pooled Over All Stations During East Winds

A second study pooling all of the data was performed using only the data collected when the wind was within plus or minus 45 degrees of an east wind. Since the wind direction is limited, the difficulty created by the selection of the wind vector is less important. Table 4 shows that even though it was necessary to reduce the dimensionality to 20 dimensions because we now only have 181 learning cases, the performance of the algorithm was improved giving a correlation coefficient of 0.65 which corresponds to explaining 25% of the variation. Figure 7 presents the relative importance vector for this algorithm. Comparing this to Figure 6, we see that variables 10-19 are all positive and thus the absolute magnitude of the wind is now playing a significant roll. It is interesting to note that the effect in the spread in the previous day's wind, variables 73 and 75, is now opposite to that seen in Figure 6. One explanation for this is that when the magnitude information for the current day's wind is available the algorithm no longer trys to estimate the current day's wind from the spread in the previous day's wind. Table 5 summarizes the most important variables indicated by Figure 7. The spread in the previous day's wind is still the most important variable although the next most important variable has now become the magnitude of the current day's wind followed by the presence of white caps.

In order to determine if the selection of the wind vector still effects the performance when the winds were restricted to within plus or minus 45 degrees of an east wind, an east wind algorithm was rederived modifying the data vectors such that the absolute values of the projection of the wind on the position directions and normal to the position direction were used. Although this eliminates the effect of wind direction, it now allows the linear regression algorithm to use the wind magnitude even when the data are pooled over all of the stations. Since the wind direction is limited to within plus or minus 45 degrees of the east wind, this is not a severe restriction. Table 4 shows that the algorithm based on the pooled absolute wind shows significant improvement in performance with a correlation coefficient of 0.78 and an explanation of 38% of the variation. This algorithm is presented in Appendix D. The relative importance vector showing the effect of each of the environmental variables shown in Table 1 on the estimated salt concentration is shown in Figure 8 and the most important of these variables summarized in Table 6. Comparison of Figures 7 and 8 verifies that the form of the wind data vectors selected still has a significant effect even when the wind is restricted to plus or minus 45 degrees, but a significant amount of the detrimental effect may be eliminated by using the absolute magnitude of the wind vector. Note, that variables 10 through 19 are now significantly more important and also variables 20 through 29 have become far more important. Referring to Table 6, we see that the most important variables are now the two components of the present day's wind followed again by the spread of variation in the previous day's wind and the presence of white caps. Note, that some of the variables associated with relative humidity and the position of the station in the east-west direction are approximately as important as these latter variables. Thus, from Figure 8, we may conclude that for the case of east winds, the most important environmental factor for determining the background concentration of salt is the magnitude of the current day's wind. This is followed by the distance of the station from the ocean, the previous day's variation in wind, the presence of white caps and the humidity all of which have approximately an equal influence of about 1/10th that of the magnitude of the wind. To obtain similar conclusions on the effect of environmental variables for all of the wind directions it would be necessary to redefine the wind vector by projecting it on two perpendicular compass directions and then rederiving both the optimal base and the regression algorithm.

BACKGROUND CONCENTRATIONS AT EACH STATIO

Regression algorithms were developed to estimate the background concentrations at each of the individual measurement stations. These algorithms have the advantage that the variation due to position need not be explained by the regression algorithm and that the optimal definition of the wind vector for estimating the concentration with the devices operating is also optimal for estimating the background conditions. Thus, these algorithms were used to estimate the background concentration which would occur at the times when the downwind data were taken. The second and third columns of Table 4 summarize the number of learning cases which were available for developing each

FIG- 7 RELATIVE IMPORTANCE VECTOR FOR EAST WIND (45-N=20)

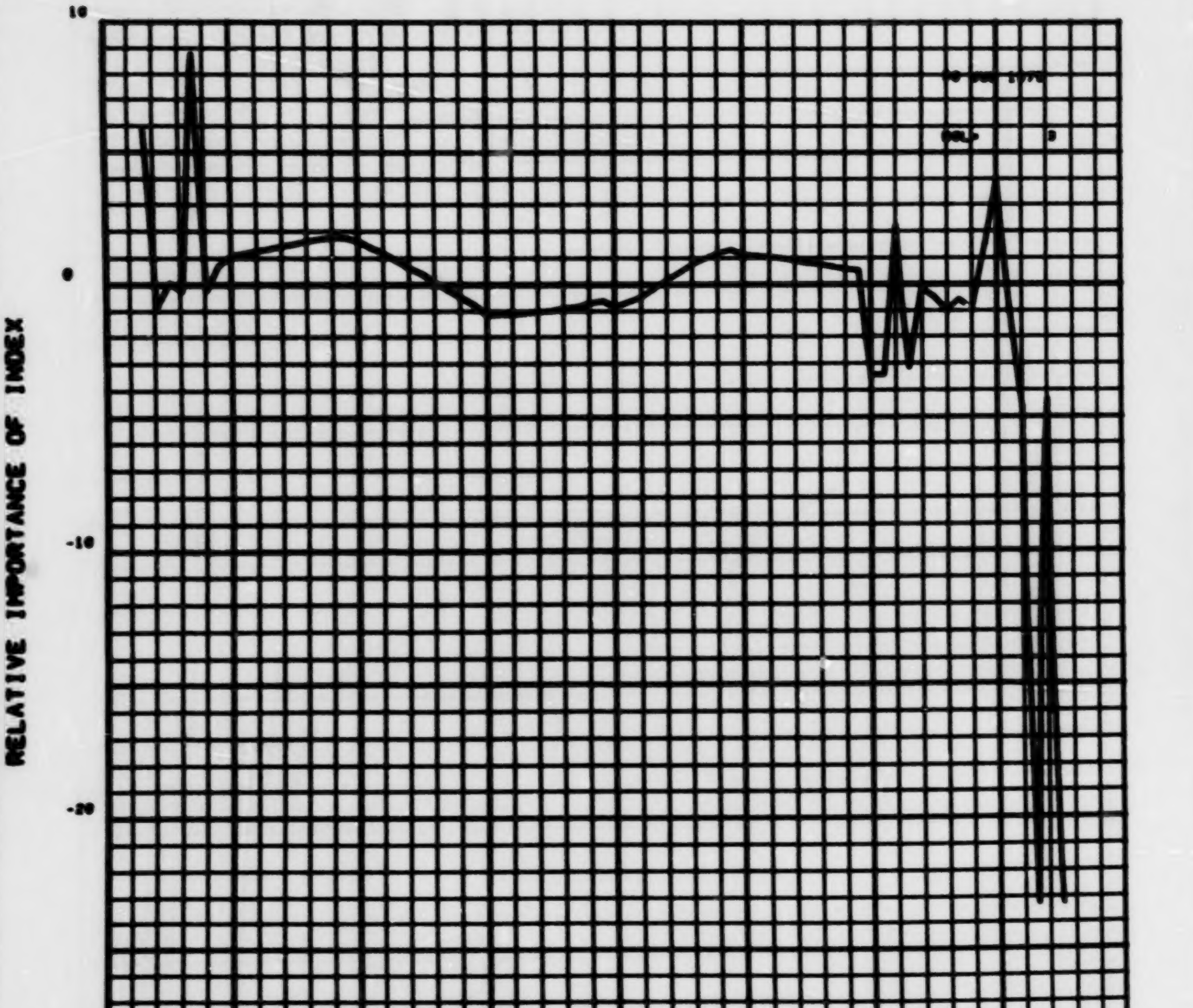


FIG-8 RELATIVE IMPORTANCE VECTOR FOR ABSOLUTE EAST WIND (45-N=20)

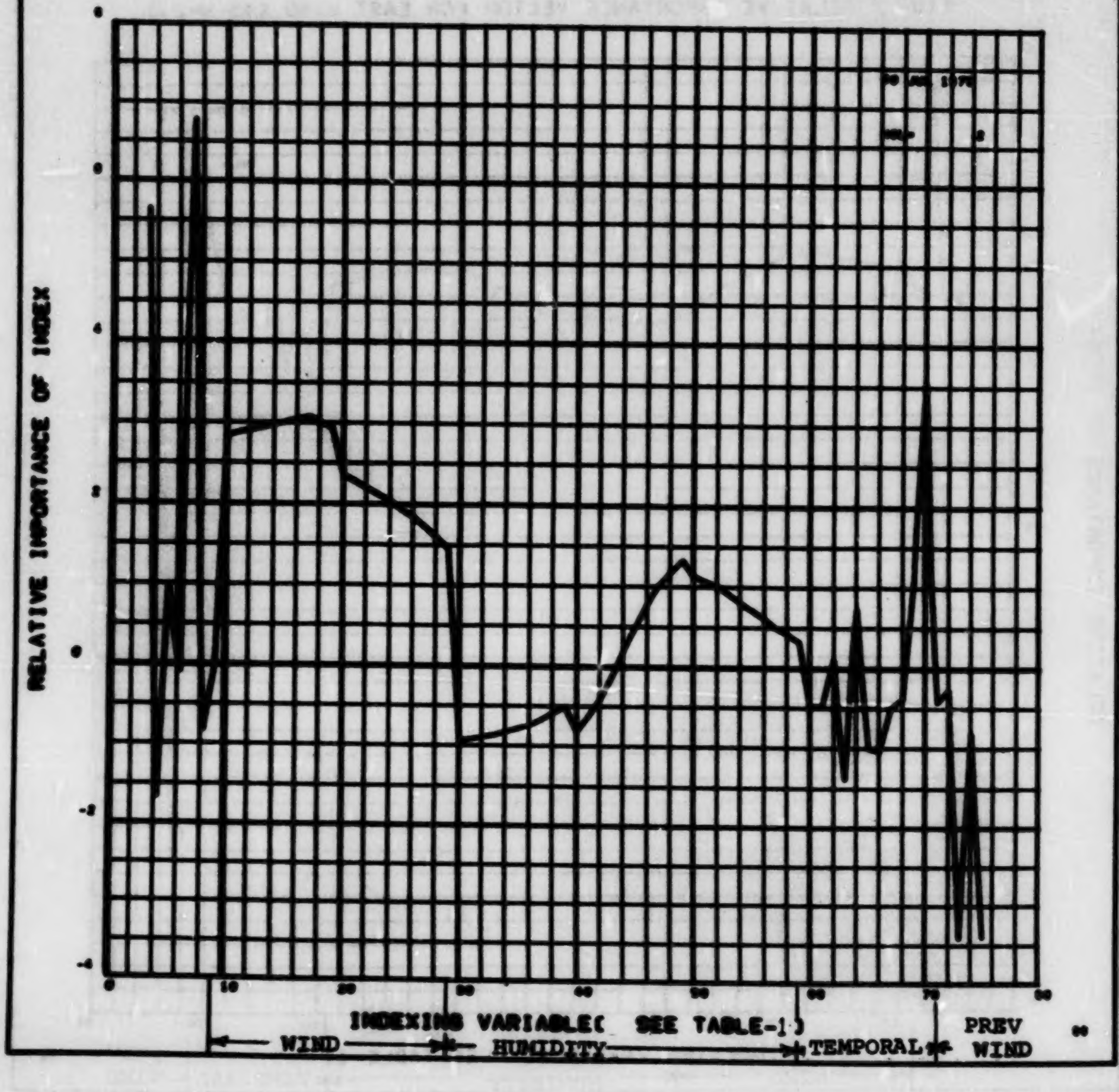


TABLE 6 - MOST IMPORTANCE ENVIRONMENT VARIABLES FOR ESTIMATE OF BACKGROUND SALT CONCENTRATION (BASED ON ABSOLUTE EAST WIND DATA)

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON ESTIMATE INCREASE	DECREASE
WIND SPEED DURING THE TEST	10-29	400	X	
WHITE CAPS	7	6.8	X	
HUMIDITY	50-59	6	X	
DRY BULB TEMPERATURE	30-39	6		X
DISTANCE OF STATION FROM THE OCEAN	1	5.7		X
PRECEDING DAYS SPREAD AND STD DEV IN WIND SPEED	73,75	7		X

TABLE 7 - MOST IMPORTANCE ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATIONS 1 AND 2

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON ESTIMATE INCREASE	DECREASE
SPREAD AND STANDARD DEVIATION OF PREVIOUS DAYS WIND	73,75	150	X	
HUMIDITY	50,59	40	X	
TESTS PERFORMED IN THE SPRING	62	27	X	
DIFFERENCE BETWEEN DRY AND WET BULB TEMPERATURE	40-49	25	X	
LIGHT RAIN	4	11	X	
PROJECTION OF WIND ON POSITION DIRECTION	10-19	10	X	
END TIME	9	9	X	
START TIME	8	8		X

of these algorithms and the number of dimensions used. Performance of the algorithms for each of these stations is summarized in Columns 4 and 5 of Table 4. These algorithms are included in Appendix D. The performance ranges from very poor at Station 11 to marginally good at Stations 1, 2 and 10. The last four columns of this table show the measured mean and standard deviation of the background concentration at each of the stations as well as the predicted background mean and standard deviation of the concentrations which would have occurred at each of the stations during the time period when the cooling devices were operating. Comparison of the mean concentrations for the learning data (i.e. the time period which the cooling devices were not operating) and the mean obtained when the cooling devices were operating shows that they are usually within a standard deviation of each other. The standard deviation of the estimates is smaller than the measured cases. This is due to the elimination of noise from the estimated values, which reduces the spread about their mean value.

Examination of Table 4 shows that, in general, the explained variation for the algorithms developed for the independent stations is less than the explained variation for the algorithm developed for the stations pooled over the absolute east wind. There are two reasons for this lower value of the explained variation. The first is that there is less variation to explain since the algorithm pooled over all of the stations must also explain the variation due to station location. If this variation is relatively easy to explain, the percentage of variation explained would be greater when this variation is included in the data. The second reason is that the number of learning cases is significantly reduced for the individual stations and, as discussed in Appendices A and C, this reduces the number of dimensions which can be used for the analysis. The performance of the algorithms developed for those stations for which relatively large number of cases were available is in general not too different from that obtained for the cases which were pooled over stations with the absolute east wind. This implies that the use of a new base with the redefined wind vector with the data pooled over all of the stations would not yield a significantly better estimate for the concentration. However, the advantage of this new pooled algorithm would be a capability to predict the background concentration at almost any point in the vicinity of the Turkey Point facility.

The calculated background concentration at each of the stations could be improved either by deriving a separate base for each of the stations, thus eliminating the position variation from the ADAPT base and decreasing the number of dimensions required to achieve a given degree of representation, or by increasing the number of learning cases available or a combination of the two. Since only a relatively small improvement will occur as a result of developing a new base at each station and approximately a factor of two increase in the dimensionality is required, it is recommended that for any future test at least 250 to 300 measurements be made at each station. It should be noted that the improvement in the estimate of the background conditions will decrease the standard deviation of the estimate, but the mean value of the background concentration averaged over all of the stations will probably remain approximately the same as was found in this study. It is unlikely that any improvement in the ability to estimate the background concentration will allow a better estimate of the effect of the cooling device unless there is a corresponding improvement in the ability to measure the concentration.

Tables 7 through 16 summarize the relative importance of the environmental variables for the concentration at each of the stations. These tables summarize the corresponding relative importance plots such as those presented in Figures 9 and 10. The relative importance vectors for the other stations are included in Appendix C. It is interesting to compare the relative importance plot presented in Figure 9 for Station 10 with the relative importance plot presented in Figure 8 for the pooled stations with an east wind. The dominate variable, the magnitude of the wind, is still the same. However, the effect of the station location Variable No. 1 in Figure 8 is now very nearly 0. This should be the case since the location of Station 10 is fixed and this algorithm does not include the variation in the position in learning data. We also see that the effect of white caps and the spread in the previous day's wind is not important to Station 10. Examination of Table 7 shows that these variables are important in Stations 1 and 2. Thus, we conclude that the effect of the previous day's wind is most important for those stations located near the ocean and becomes relatively unimportant for those stations located far back from the ocean. Figure 9 shows that Variable 64, which Table 1 shows as a test performed in the fall is more important then was observed in Figure 8. Examination of Tables 7

TABLE 8 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 3

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE	
			INCREASE	DECREASE
Projection of Wind on Position Direction	10-19	20	X	
Dry Bulb Temperature	30-39	10		X
Variation in Relative Humidity During Test	40-59			
Projection of Wind Vector on Normal to Position Direction	20-29	6		X
Tests Performed in the Fall	64	3	X	

TABLE 9 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 4

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE	
			INCREASE	DECREASE
Projection of Wind Vector on Position Direction	10-19	15		X
Projection of Wind Vector on Normal to Position Direction	20-29	9	X	
Relative Humidity	50-59	9		X
Difference Between Dry & Wet Bulb Temperature	40-49	6	X	
Dry Bulb Temperature	30-39	5		X
Tests Performed in the Summer	63	3		X

TABLE 10 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 5

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE	
			INCREASE	DECREASE
Projection of Wind on Position Direction	10-19	10		X
Projection of Wind on Normal to Position Direction	20-29	5	X	
Dry Bulb Temperature	30-39	4		X
Variation in Humidity	40-59			
Tests Performed in the Summer	63	2½		X
Tests Performed on Friday	69	2	X	

TABLE 11 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 6

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE	
			INCREASE	DECREASE
Projection of Wind Vector on Normal to Position Direction	20-29	24	X	
Projection of Wind Vector on Position Direction	10-19	18		X
Dry Bulb Temperature	30-39	10		X
Variation in Humidity	40-59			
White Caps	7	5	X	

TABLE 12 - MOST IMPORTANT ENVIRONMENT VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 7

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE INCREASE	EFFECT ON PRE DECREASE
Projection of Wind on Position Direction	10-19	20	X	
Projection of Wind on Normal to Position Direction	20-29	20	X	
Tests Performed in the Fall	64	13	X	
Tests Performed in the Summer	63	8		X
Variation in Humidity	40-59			

TABLE 13 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 8

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE HUMIDITY	EFFECT ON PRE INCREASE	EFFECT ON PRE DECREASE
Projection of Wind on Position Vector	10-19	22		X
Difference Between Dry and Wet Bulb Temperature	40-49	5	X	
Dry Bulb Temperature	30-39	2	X	
Relative Humidity	50-59	2		X
Days Since First Test	60,61	3		X

TABLE 14 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 9

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE HUMIDITY	EFFECT ON ESTIMATE INCREASE	DECREASE
Magnitude of the Cross Wind	20-29	21	X	
Magnitude of Wind Towards Cooling Device	10-19	5.5		X
Difference Between Dry and Wet Bulb Temp	40-49	2		X
Humidity	50-59	2	X	
Test Performed in the Spring	62	2		X
Test Performed on Friday	69	1.2		X
Test Performed in the Fall	64	1.1	X	
Test Performed on Wed	67	1.1	X	

TABLE 15 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 10

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE HUMIDITY	EFFECT ON ESTIMATE INCREASE	DECREASE
Magnitude of Wind On Position Direction	10-19	28	X	
Wind Magnitude Normal to Position Direction	20-29	17	X	
Difference Between Dry and Wet Bulb Temp	30-39	7		X
Test Performed in the Fall	64	4.5		X
Length Time Since First Test	60,61	2	X	

TABLE 16 MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF BACKGROUND CONCENTRATION AT STATION 11

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE IMPORTANCE	EFFECT ON PRE INCREASE	EFFECT ON PRE DECREASE
Projection of Wind Vector on Position Direction	10-19	11		X
Difference Between Dry and Wet Bulb Temperature	40-49	5	X	
Dry Bulb Temperature	30-39	4		X
Relative Humidity	50-59	3		X
Tests Performed in the Spring	62	3		X
Days Since First Test	60,61	4		X
Projection of Wind on Normal to Position Direction	20-29	1.5		X
Tests Performed on Monday	65	1½	X	

FIG-9 RELATIVE IMPORTANCE VECTOR FOR BACKGROUND CONC AT STATION-10 N=12

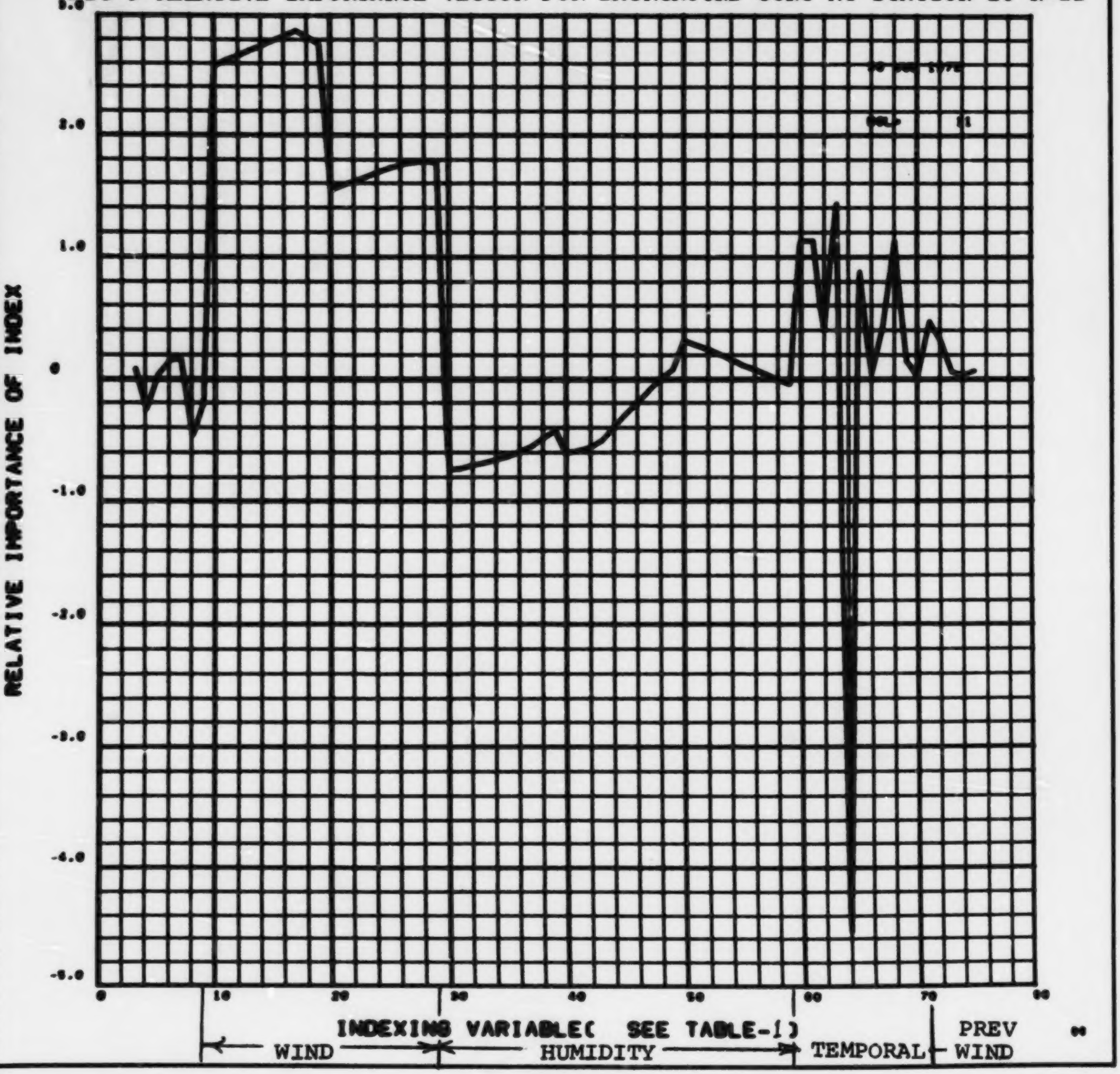
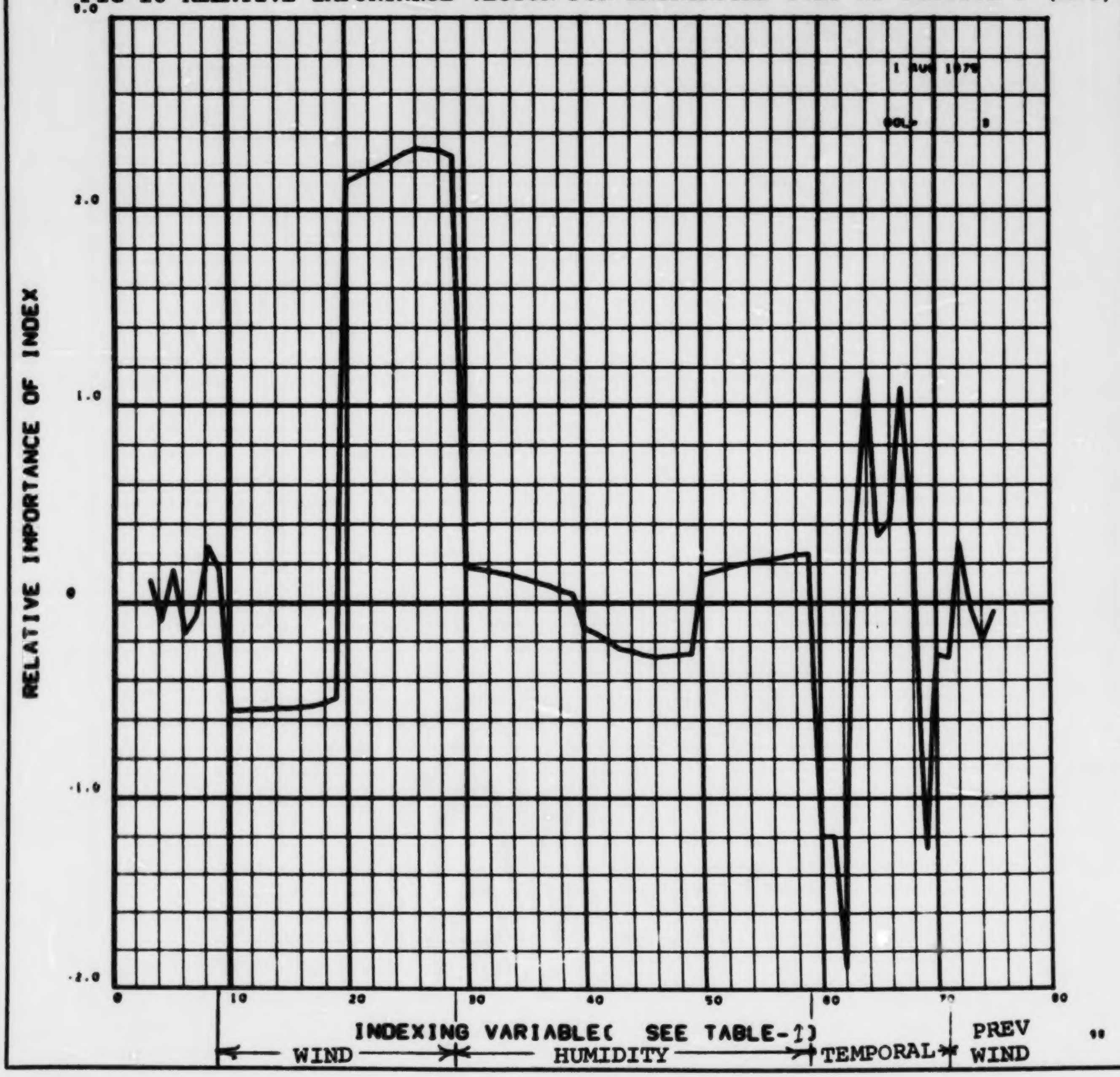


FIG-10 RELATIVE IMPORTANCE VECTOR FOR BACKGROUND CONC AT STATION-9 (N=8)



through 16 and the corresponding figures in Appendix 3 show that the seasonal effects vary considerably from station to station. Thus, a pooled estimate over all stations tends to lose these seasonal effects. The reason why the seasonal effects are different from station to station is unknown.

Comparison of Figure 10 with Figure 9 verifies the previous explanation for why the estimate pooled over all stations without limitation to wind direction loses the importance of the magnitude of the wind. At Station 9 the effect of the wind projected on the position direction is exactly opposite to that for Station 10. Thus, when pooled over all stations, the data for these two stations tends to cancel the effect of wind direction.

Examination of the relative importance of the wind for each of the stations also shows the effect of distance of the station from the ocean and the cooling basin located South of the cooling devices. For Stations 1 and 2 the dominate effect is the variability of the previous day's wind and the current day's wind is not as significant. This suggests that for stations very near the ocean, the most important factors are those controlling the salt concentration over the ocean. In going from Station 3 to Station 4, the reader should recall that the definition of the wind vector is such that the sign of the effect of a given wind is reversed because of the change in coordinate system. Thus, at Station 3 if the projection of the wind on the position direction is positive at Station 4 the same wind projection is negative. At Stations 3, 4 and 5 all of which are located on an East-West line adjacent to the cooling basin, the ratio of the effect of the projection of the wind direction on the position direction (i.e. the East-West wind component) is of the same order as the wind component in the North-South direction. However, for Station 6 which is located on the same East-West line but not directly adjacent to the cooling basins, the effect of the North-South wind relative to the East-West wind has been increased. Thus, as we move further away from the cooling basins, we see a similar affect to that which was observed as we moved away from the ocean. This suggests that the cooling basins probably have a significant effect on the concentration at those stations located close to them. Thus, it is recommended that for any future experiments any open water, including cooling basins, be considered as sources of background salt concentration.

EFFECT OF COOLING TOWER ON AMBIENT CONCENTRATION

The effect of the cooling tower on the ambient concentration was obtained by using the algorithms discussed in the preceding section and summarized in Table 4 to estimate the background salt concentration which would have occurred during each of the tests where the cooling tower was operating. This calculated background concentration was subtracted from the measured concentration to determine the salt deposition which could be attributed to the cooling tower. When this difference was averaged over the entire set of 398 measurements made during the operation of the cooling tower, the average increase in ambient concentration over the expected background concentration was 0.002 micrograms per cubic meter with a standard deviation of 4.8 micrograms per cubic meter. This clearly indicates that on the average the increase in concentration was less than could be measured in this test program.

Similar analyses were made for each of the individual stations. The results of these analyses are summarized in Table 17. The first column of Table 17 indicates the station number for which the statistical summary is provided. The second column indicates the number of cases for which measurements were made at that station while the cooling tower was operating. The third column gives the mean value of the difference between the measured concentration and the expected background concentration at that station. The standard deviation of this average is provided in the fourth column. The fifth and sixth columns provide the maximum and minimum values which were observed for the difference between the measured and the expected background concentration. The seventh column gives the confidence that the increase in concentration at the station due to the cooling tower was less than the standard deviation of the measurement. The average of this confidence for all stations was 83%. This table shows that for Station 10 there were only three cases and thus a statistical summary has no meaning for this station. Thus, for the remainder of this discussion we shall only consider Stations 3-9 and 11. For each of these stations we note that the standard deviation of the measurement exceeds the mean value of the measurement. This fact alone is a strong indication that no measurable enhancement of the expected background concentration resulted from operating the cooling tower. A further strengthening of this conclusion beyond the 83% confidence level is the fact that approximately half of the mean values are negative and the other half are positive. This is exactly the situation that would be expected in the event the cooling tower had no effect upon the background

TABLE 17 - SUMMARY STATISTICS FOR DOWN WIND MINUS BACKGROUND CONCENTRATION FOR COOLING TOWER

STA	NO CASES	MEAN mgr/m^3	STD DEV ugr/m^3	MAX VALUE ugr/m^3	MIN VALUE ugr/m^3	CONFIDENCE THAT INCREASE IS LESS THAN STD DEV (%)
3	21	-0.3101E 01	0.3590E 01	0.4552E 01	-0.1115E 02	97
4	53	0.4646E 01	0.5593E 01	0.3181E 02	-0.1520E 01	57
5	54	-0.3211E-01	0.4238E 01	0.1637E 02	-0.6419E 01	84
6	59	-0.2417E 01	0.3095E 01	0.5779E 01	-0.7797E 01	96
7	51	0.2553E 01	0.3611E 01	0.1170E 02	-0.6590E 01	61
8	63	-0.1976E 01	0.3975E 01	0.1022E 02	-0.8790E 01	93
9	50	-0.1236E 01	0.3349E 01	0.1024E 02	-0.7463E 01	92
10	3	-0.4546E 01	0.1170E 01	-0.2900E 01	-0.5504E 01	-
11	44	0.7658E 00	0.4658E 01	0.1560E 02	-0.6409E 01	85

TABLE 18- SUMMARY STATISTICS FOR DOWN WIND MINUS BACKGROUND CONCENTRATION FOR SPRAY MODULES

STA	NO CASES	MEAN ugr/m^3	STD DEV ugr/m^3	MAX VALUE ugr/m^3	MIN VALUE ugr/m^3	CONFIDENCE THAT INCREASE IS LESS THAN STD DEV (%)
3	21	-0.1451E 00	0.7726E 01	0.1799E 02	-0.8405E 01	84
4	41	0.1166E 02	0.1997E 02	0.1036E 03	-0.7852E 01	66
5	43	0.7576E 00	0.6408E 01	0.3179E 02	-0.5484E 01	81
6	43	-0.2935E 01	0.3166E 01	0.7399E 01	-0.1138E 02	97
7	37	0.3349E 01	0.2953E 01	0.1342E 02	-0.1011E 01	(44)
8	54	-0.2797E 01	0.4178E 01	0.1336E 02	-0.1173E 02	95
9	40	-0.1072E 01	0.2370E 01	0.8539E 01	-0.4251E 01	93
10	6	-0.5033E 01	0.1245E 01	-0.3003E 01	-0.6554E 01	-
11	40	0.3483E 01	0.9393E 01	0.4857E 02	-0.6938E 01	73

concentration. This combined with the extremely small value of the average difference over all 398 cases suggests that the increase in the salt concentration over the expected background concentration as a result of the operation of the cooling tower is probably significantly less than the measurement accuracies of the instrumentation used in this program.

EFFECT OF SPRAY MODULES ON AMBIENT CONCENTRATION

The effect of the spray modules on the ambient concentration was obtained in exactly the same way as the effect of the cooling tower. That is the algorithms discussed in Section 3.2 were used to calculate the expected background salt concentration for each of the tests performed with the spray modules operating. The average increase in the expected background concentration overall 325 tests was 1.32 micrograms per cubic meter with a standard deviation of 9.9 micrograms per cubic meter. Again we conclude that on the average the effect is smaller than could be measured by this instrumentation. Table 18 presents the summary statistics for each one of the stations with a spray module operating. This table is of the same form as Table17. Examination of this table also shows that Station 10, an inadequate number of cases were available for one to draw a meaningful statistical conclusion. For the remaining stations, we again note that approximately half of the stations had negative mean values and half had positive mean values. However, the fact that the average over all 325 was 1.32 micrograms per cubic meter shows that in general those with positive mean values had slightly larger positive mean values then those with negative. This would be an indication that there maybe some stations for which the concentration was increased by a small amount by the operation of the spray module.

Table 18 shows that for Station 7 the mean value of the difference between the observed and expected background concentration is slightly greater than the standard deviation. Since approximately 85% of the Gaussian distribution falls between plus or minus one standard deviation of the mean, there is approximately 85% confidence that at Station 7 there was some increase in the background salt concentration due to the operation of the spray modules. At the other stations, the mean value for the difference between the measured concentration and the expected background concentration is less than the standard deviation of the measuremen with an average confidence of 84%. Thus, for all of the other stations, we can conclude that the effect of the spray modules on the background concentration is smaller than the ability of the present instrumentation to measure this effect.

Station 7 represents an unusual station for this study in that it is the only station which is both relatively near the spray modules and has a small standard deviation for the difference between the measured salt concentration and the expected background concentration. This implies that if the measurements at some of the other nearby stations such as Stations 3, 4, 11 and 5 had standard deviations of the order of three micrograms per cubic meter, increases may also have been observed at these staions. Since the smallest standard deviations occurred at Stations 6, 7 and 9 which are not located adjacent to the cooling basin, one might speculate that the cooling basin itself introduced a significant amount of variation into the measurements. In this case, it is suggested that the location of Station 7 away from the cooling basins may have been the reason why the accuracy was sufficient that one was able to measure an increase due to the effect of the spray modules. This leads to the recommendation that for future tests one should locate the majority of the measurement stations at least 300 meters from the cooling basin. We note that an increase in the error due to the proximity to the cooling basin is in general agreement with the results obtained by examination of the relative importance vector for the background concentration which indicated that the cooling basin does have a significant effect on the salt concentration and is in part strongly dependent on the magnitude and direction of the wind. It is not unreasonable to assume that there may be other factors which affect the increase in salt concentration due to the cooling basin which have not been included in the present study.

Since the results in general for the effect of the cooling devices on the background concentration were negative, algorithms for estimating this effect could not be prepared except for Station 7. Thus, it is impossible to use the present data to obtain algorithms for estimating the effect of position and environmental variables on the increase in the background concentration due to the operation of either the cooling tower or the spray modules. However, the data from Station 7 offer a potential for making an algorithm to determine the effect of the environmental variables on the salt concentration at the location of Station 7. Since the standard deviations are small, the data from Station 6, 7, 9 and 10 may be pooled to obtain a limited effect of position on the salt concentration. Thus, algorithms were developed using the data from Station 7 to predict the different between the measured and expected background salt concentration for the spray module. A similar algorithm was developed using the data for Station 6 through 9. The performance of these algorithms were quite poor with correlation coefficients of 0.40 and 0.60 for the Station 7 and Station 6 through 9 algorithms, respectively. The fact that the correlation coefficient is greater for the Station 6 through 9 algorithms is the result of the fact that a larger number of cases were available to make this algorithm. Thus, a dimensionality of 16 could be used which allowed one to incorporate a significant portion of the information using the base derived for the anlaysis of this salt spray data. For Station 7, there were only 37 cases available and thus the maximum dimensionality which could be used was 4. This corresponded to only using approximately half of the information in the data. Thus, the performance of this algorithm is severely restricted by the limited number of cases available for the analysis. This restriction could be alleviate to some extent by the development of a new base using only the data from Station 7 for the Station 7 algorithm and only the data from Station 6 through 9 to develop this Station 6 through 9 algorithm. Since this new base would not have to incorporate the information regarding position variation or the information concerned with the other stations it would be a better base for the analysis of the data from these limited subsets of stations.

Although the performance of the algorithms for Stations 6-9 was very poor, there is reason to believe that the dominate variables occurring in the relative importance vectors will probably remain important even if more data were available for the analysis. Thus, the relative importance vectors for these two algorithms are presented in Figures 11 and 12. Figure 11 presents the relative importance vector for the algorithm for estimating the increase in ambient salt concentration due to the spray module at Station 7. The dominate variables for this algorithm are summarized in Table 19. Similar results are presented for Stations 6 through 9 in Figure 12 and summarized in Table 20. In both of these cases, only a few of the most dominate variables are included in these tables because of the

FIG 11 RELATIVE IMPORTANCE VECTOR FOR SPRAY MODULE STA 7 (N=1)

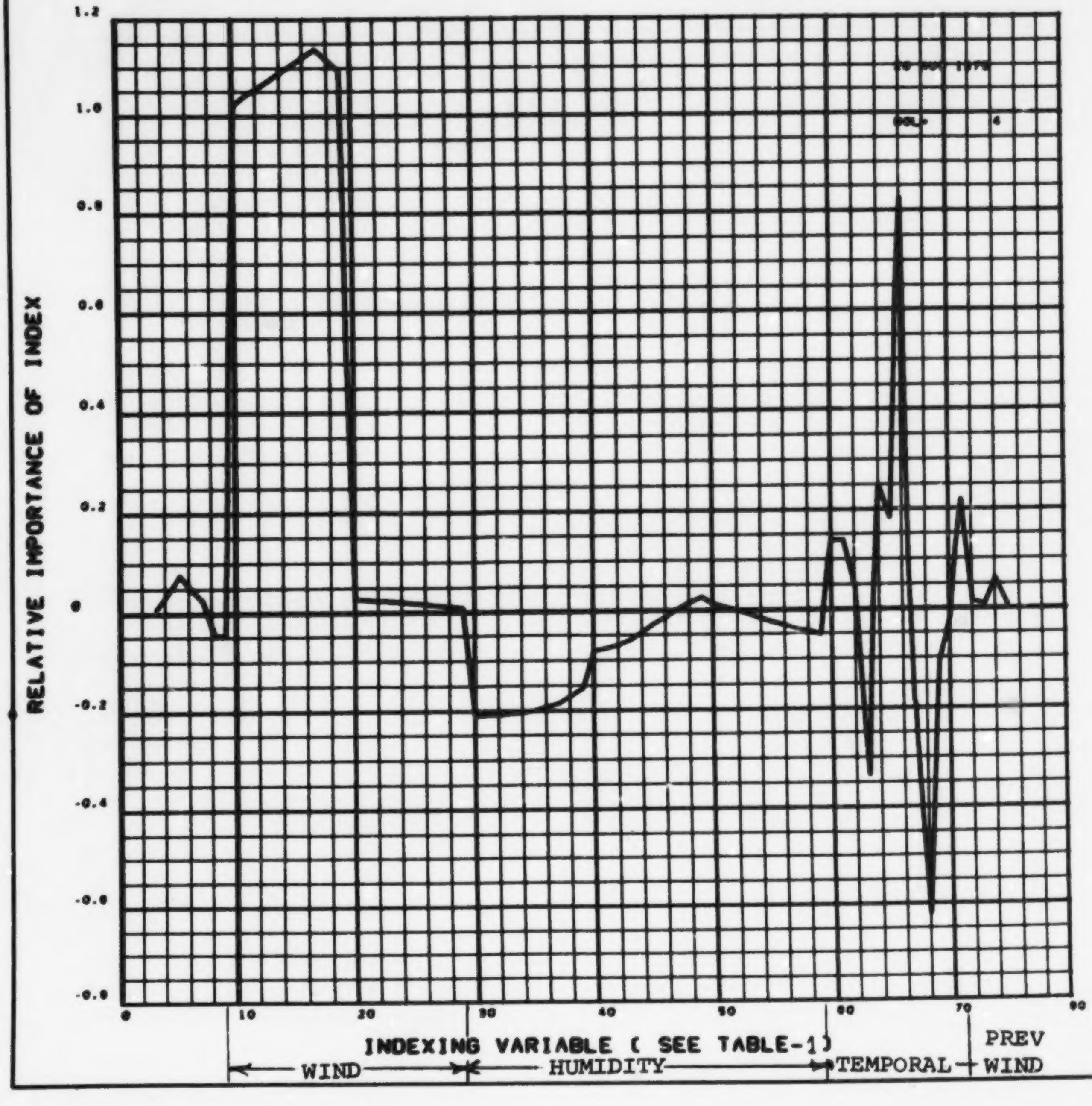


FIG-12 RELATIVE IMPORTANCE VECTOR FOR SPRAY MODULE STAG-9 (N=16)

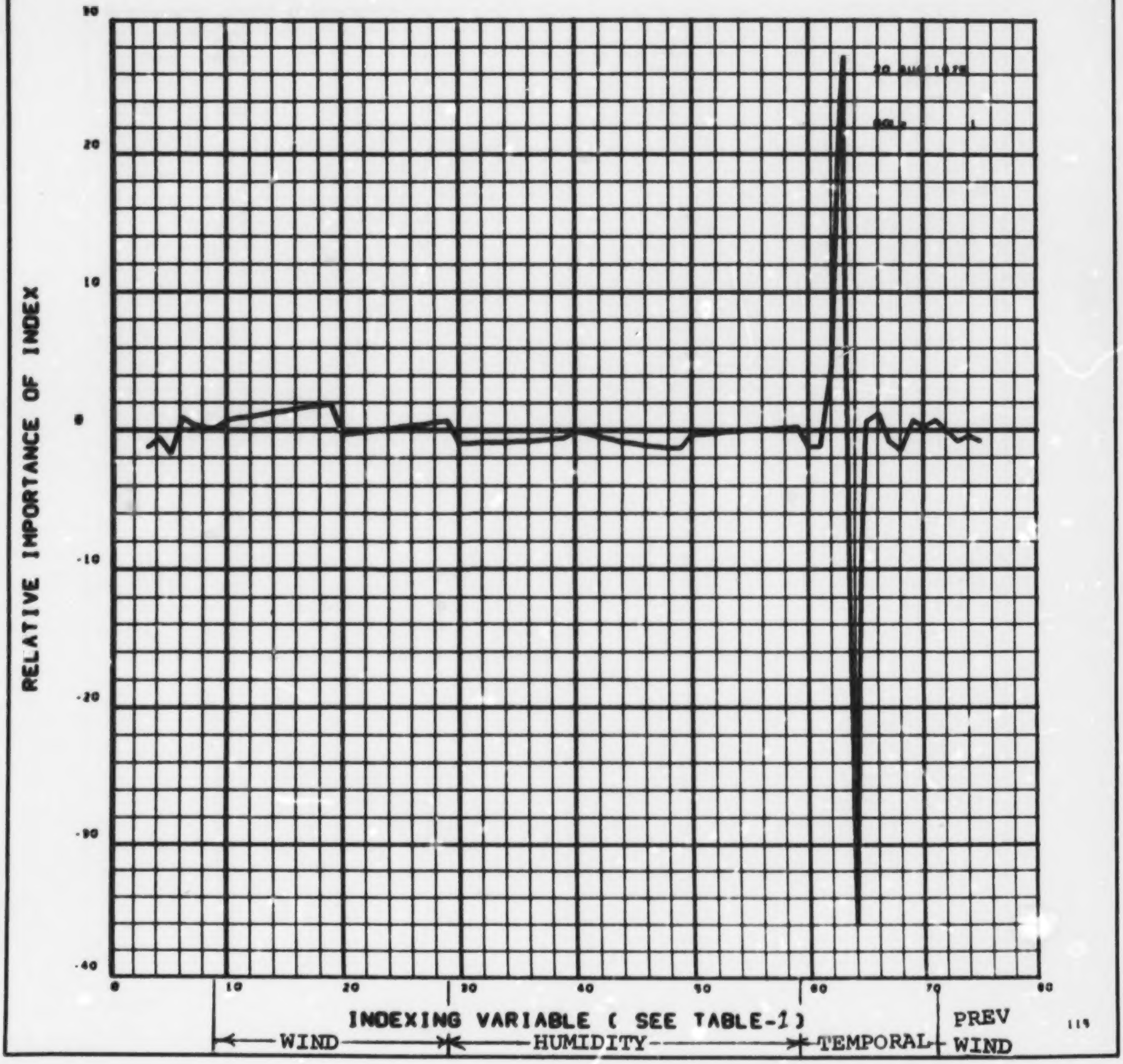


TABLE 19 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF INCREASE IN SALT CONCENTRATION DUE TO SPRAY MODULES AT STATION 7

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE HUMIDITY	EFFECT ON ESTIMATE INCREASE	DECREASE
Projection of Wind on Position Direction	10-19	16	X	
Projection of Wind on Normal to Position Direction	20-29	2	X	

TABLE 20 - MOST IMPORTANT ENVIRONMENTAL VARIABLES FOR ESTIMATE OF INCREASE IN SALT CONCENTRATION DUE TO SPRAY MODULES AT STATIONS 6 THROUGH 9

ENVIRONMENTAL VARIABLE NAME	INDEX NO.	RELATIVE HUMIDITY	EFFECT ON ESTIMATE INCREASE	DECREASE
Test Performed in the Fall	64	96		X
Test Performed in the Summer	63	27	X	
Projection of Wind on Position Direction	10-19	20	X	

poor performance of these algorithms which suggests that only the most significant of the variables can be considered meaningful. It is recommended that if additional information is desired regarding the amount of increase at Station 7 as a function of variation in the environment is desired that these data be reprocessed through the ADAPT programs to derive an optimal base for this station by itself and that this base be used to rederive an algorithm to estimate the effect of the environment on the concentration of Station 7. A similar procedure is recommended for the data obtained from Stations 6-10. This will result in significantly improved relative importance vectors and possibly an algorithm allowing application of the results of this study to other power plant sites. (See Section 4.0)

SECTION IV

EXTENSION TO OTHER SITES

The approach to the present study consisted of determining the effect of the cooling devices on the ambient concentration by subtracting the background concentration expected at the Turkey Point site. Thus, any measured increase in concentration found in this analysis would be independent of the site at which these specific cooling devices are located. However, since the effect observed was essentially no increase, these results are trivial except for the effect of the spray module at Station 7. Thus, we may state in general that the effect of the particular cooling tower used in the present study would be smaller than the accuracy of the present measurements regardless of the site at which it was located. We may also state, based on the results of the observations of Station 7, a pair of spray modules as used in this study can be expected to increase the background salt concentration by approximately three micrograms per cubic meter at distances of approximately 400 meters from the spray module averaged over the environmental conditions similar to those observed during the present test program.

Although the estimated increase in the background concentration due to the cooling devices obtained as a result of this study are independent of the site with respect to the measurement accuracies found in this study, the measurement accuracies themselves are not independent of the site. This can be seen from the fact that the most likely reason that an effect was observed at Station 7 and not at Stations 3, 4, 5 and 11 was the fact that Station 7 was located at a greater distance from the cooling basins then the other stations and thus had more accurate estimates. Thus, we must conclude that a similar study made at a different site where the measurement accuracies can be expected to be more similar to those observed on Station 7 would result in measurable increases in the background concentrations at least at the stations located near the spray module.

In the absence of the additional analysis recommended in Section 3.4 and/or testing required to develop a site independent algorithm for the spray modules, it can only be concluded that for environment conditions such as those observed at the Turkey Point site the average increase in background salt concentration due to a pair of spray modules is approximately three micrograms per cubic meter. The relative importance vectors show that these concentrations will be strongly affected by the wind and the season during which the tests are performed. Clearly, many other variables are still important and could only be defined through a more complete analysis of additional test data.

The results obtained for the estimate of the ambient concentration are peculiar to the Turkey Point location. However, the relative importance vectors allow one to generalize these results in terms of the characteristics of this site. For example, it was observed that very near the ocean, the dominant effect is the variability of the previous day's wind. This indicates that at locations near the ocean background salt concentration is determined primarily by those environmental conditions which tend to increase the amount of salt over the ocean. As one moves to larger distances from the ocean, the current day's wind becomes more important since a transport mechanism is required to transport the salt to the measurement site. In general, the season is quite important to the background salt concentration. The humidity also appears to be quite important. However, this may be the result of the particular measurement instrumentation used. Some effects indicated by the relative importance tables represented in the previous section may also be due to the measurement procedures. This is particularly likely for some of the variables associated with the relative importance vectors for the precision run error. The importance of Monday testing is almost certainly highly dependent on either the instrumentation or the test procedure.

SECTION V

REFERENCES

1. Schrecker, Gunter O., et al "Drift Data Acquired on Mechanical Salt Water Cooling Devices", Final Report EPA Contract 68-02-1365 prepared by Environmental Systems Corporation, U. S. Environmental Protection Agency Report No. EPA-650/2-75-060, July, 1975.

SECTION VI
APPENDIX - A
MATHEMATICAL DESCRIPTION OF ADAPT

SUMMARY

The ADAPT analysis techniques consist of a family of computer programs which are capable of performing empirical analysis of any type of data. The generality of these programs follows from the character of an empirical analysis which may be considered to be made up of two separate steps. The first step of an empirical analysis is the learning or training step. In this step, data for which the answer is known is processed to determine the algorithm, i.e. rule, for obtaining the answer from the data. The second step is to apply the algorithm derived in Step-1 either to proof test data to demonstrate performance or to operational data to get the desired answer. The ADAPT programs incorporate this entire procedure with a very general input format and the capability to apply a large number of the classical empirical analysis techniques to the derivation of the algorithm. This alone makes the ADAPT programs an extremly useful tool since further programming is not required in order to develop algorithms once a given data set has been properly formatted for input to the ADAPT programs.

The unique empirical analysis capabilities of the ADAPT programs arise from preceding the classical empirical analysis techniques with an efficient representation of the data. This representation enhances the subsequent empirical analysis. It reduces the dimensionality of the problem so that the empirical analysis may be applied to considerably larger data sets. This approach has the additional benefits of requiring less learning data, providing both empirical validity criteria and additional insight to the nature of the data. The optimal representation is obtained by transformation of the data to the ordered optimal coordinate system as defined by the Karhunen-Loeve expansion (see reference 1) this optimal representation is also known as principal component analysis, optimal empirical orthogonal functions and is closely related to factor analysis. The ADAPT programs incorporate a unique approach to numerically deriving this transformation. The ADAPT programs can derive this transformation for an unlimited number of vectors having over 2,000 components each. This capability represents an order of magnitude increase over what can be accomplished with the classical approach to finding the Karhunen-Loeve expansion.

Representation in this optimal space usually requires only 1/10th to 1/100th as many numbers as was required by the original format of the data. With the data represented efficiently, any of the classicial empirical techniques including regression, classification, pattern recognition and clustering may be performed with considerably less computational resources and also with a smaller amount of learning data. The detailed description of the methods used to obtain the optimal representation and the use of this optimal representation to improve empirical data analysis will be presented in four parts: 1) definition of the data histories, 2) description of the optimal representation of the data histories, 3) description of the use of the optimal representations for empirical data analysis and 4) evaluation of algorithm performance.

DEFINITION OF DATA HISTORIES

Empirical analysis in general and the ADAPT techniques in particular address themselves to the analysis of information which appears as data histories, where data histories are defined as an indexed series of numbers which convey information. Although the indexing variable is often time or some other continuous function, it can be anything. The histories may consist of numbers with different physical meanings. For example, ADAPT analyses have been performed on data histories consisting of pressure versus time adjourned to dimensional measurements associated with the hardware which produced the pressure versus time history. ADAPT analyzes have also been performed on data consisting of temperature as a function of spacial location adjoined to quanities such as latitude, longitude and day of the year.

The histories may be given in continuous (analog) form or in descreet form; since the ADAPT programs operate in digital computers, analog histories are each digitized into a set of N numbers. Thus, a history is treated as an N dimensional vector in Euclidean space. If there are N histories the result is an M by N matrix of numbers which represent a given ensemble of information.

It may be desirable to perform some pre-processing on any given data set to bring out features or characteristics of this data before entering the ADAPT programs. Such pre-processing can be performed using the ADAPT programs and include such non linear pre-processing as normalization, raising to a power, taking logarithms or anti-logarithms, taking Fourier transforms, equalizing the data, etc. The particular pre-processing which is required for any given problem is normally suggested by previous data processing experience and apriori knowledge of some significant characteristics of the data. For example, data containing a large number of irrelevant spikes should be pre-processed by taking the log of the data. On the other hand, if the spikes contain the significant information the anti-log of the data might be more useful. If the data consists in part of quanities which are totally unrelated and therefore measured by different units, the relative magnitude of two such quanities such as temperature and length is entirely dependent on which units are selected; for example, if degrees and miles are chosen as the units, the ratio of the magnitudes of the distance measurements to the temperature measurements is considerably smaller than if degrees and angstroms are selected as the units. To compensate for this the ADAPT programs have the capability of introducing an equalization which emphasizes the variation in the data rather than the absolute magnitude. This equalization is accomplished by adjusting the magnitude of each variable, V, by the following law:

$$V_{eq} = 1 + \frac{V - MIN}{MAX - MIN} \qquad (1)$$

where MAX and MIN are the largest and smallest values of V which occur . When each variable is processed by this law it has a maximum value of two, a minimum value of one and all other values fall between one and two. Normalizations based on producing all data vectors with unity absolute magnitude and normalization based on the first data history are also available in the ADAPT programs.

OPTIMUM REPRESENTATION OF DATA HISTORIES

The choice of the N numbers which were used to represent each data history, is to some extent arbitrary, the chief criterion being that the desired physical phenomena are properly contained in the N numbers. From a theoretical viewpoint, one could use a continuous data history, however, this would require processing of an infinite number of numbers. Clearly, the realities of numerical analysis on digital computers require that the input be in vector (digitized) form, rather than functional (analog) form. Thus, the first problem is the classical sampling problem. In addition to this problem, it is possible through proper choice of coordinate transformations to still further reduce the number of numbers required to represent a given amount of information. The approach taken in the ADAPT programs is to solve these two problems sequentially. Thus, the first step in the optimization is to optimize the sampling of the data matrix. The second step in the optimization is to find the best coordinate system for representing the data. These two steps will now be discussed.

ADAPT SAMPLING PROCEDURE

The first step in finding the optimal base for the ADAPT programs is to examine the entire ensemble of data to determine how to best sample the data matrix. Here best is defined as that sampling which contains more than a specified amount of new information as defined by the sampling criteria. The results of this procedure may be considered equivalent to an incomplete classical Gram-Schmidt procedure. The degree of incompleteness is a function of the sampling criteria. The resulting dimensionality of the new orthogonal represetation is also a function of the sampling criteria. The trade-off between the two conflicting requirements of maximizing the completeness of the base and minimizing the dimensionality of the new representation is accomplished by varying the sampling criteria. The impact of the degree of incompleteness of this first step on the final representation will be discussed in Section 3.3. For the special case of selection criteria of the order of unity this first step in the ADAPT optimization reduces identically to that of the Gram-Schmidt procedure.

Since the first step of the optimization procedure is not a classical technique, the best way for the reader to comprehend the result of this step is to consider it as a modification of the Gram-Schmidt procedure. The process retains the capability of the Gram-Schmidt procedure to discreetize the data regardless of the form of the input or of its dimensionality. That is, the Gram-Schmidt base vectors obtained from a set of data histories represent these histories with a discreet number, NC, of components even when these data histories are continuous functions. Since the Gram-Schmidt procedure eliminates all linearly dependent cases, the value of NC must be less than the number of cases which are processed through the Gram-Schmidt procedure. In general, the number will be smaller yet because of linear dependance within the data set. With the ADAPT modification of this procedure, the resulting number of components is normally considerably less than even those which would result from the application of the Gram-Schmidt procedure. Therefore, after the first step of the representation the new representation already significantly reduced the dimensionality and is largely independent of the particular way the data histories were digitized.

However, just as in the case of the Gram-Schmidt base there is no reason to believe that a given representation obtained by this procedure is the best one for representing the data. Base vectors are to a great extent determined by the order in which histories were arranged when processed through either the ADAPT selection procedure or the Gram-Schmidt procedure. The next step of the ADAPT optimization is to find the base which is the best representation for the given data. This is accomplished by the second step of the optimization.

SECOND STEP OPTIMIZATION

To find the best representation a new set of NC N-dimensional orthonormal vectors, rotated from the data represented in the first step orthogonal base is postulated. This set is to be chosen in an ordered fashion, so that the first vector is the best, and so on. Only a limited number, $NR \leq NC$, of these vectors will be used as new base vectors for representing the histories. They are chosen as follows: Each history vector is represented by its coefficients in the first step base, and is projected onto the NR new vectors, giving $M \times NR$ components in the new base. If there were as many new vectors as first step vectors, $NR = NC$, and the first step selection criteria had produced a complete representation, this would be an exact representation of the history vectors. Since, in general, $NR < NC$, it is only approximate, leaving an error vector as the difference between the history vector and its representation in the new vector base, The square magnitude of this error is a measure of the error for each history, and the average of these square magnitudes for all histories is the mean square error incurred by representing the history vectors in only NR new base vectors.

The new orthonormal set of vectors is chosen by minimizing this mean square error, thus defining the meaning of a "best" set of vectors. If only one vector is used, NR = 1, it is that vector which makes the one-vector representation error the smallest. If a second vector is used also, it is chosen so that together with the first vector, it minimizes the two-vector representation error. This is continued for as many vectors, i.e., as large a value of NR < NC, as is necessary or desirable.

When formulated mathematically, this criterion requires the maximization of a quadratic form whose unknowns are the first-step components of one of the "best" base vectors, and whose coefficient matrix is the sum of the covariance matrices of the first step components of the input histories. This problem is a classical one in linear algebra, which often appears under the names of the principal components analysis of a matrix, Karhunen-Loeve or eigen function expansion and optimum empirical orthogonal function. The solutions for the unknown vector components are the normalized eigenvectors of the covariance matrix sum, and the resulting values of the quadratic form are the eigenvalues of this matrix. Once they are obtained, they are simply arranged in order of decreasing size of the eigenvalues. The largest eigenvalue gives the most reduction in mean square error that can be achieved with only one new base vector; and the corresponding eigenvector is this new base vector. The next largest eigenvalue gives the most reduction in the error that can be achieved by using a second new base vector in addition to the first one found above, and this second vector is the eigenvector of this second largest eigenvalue. This process can be continued until the desired accuracy is achieved. The sum of the NR largest eigenvalues gives the maximum mean square error reduction which can be achieved with NR new base vectors; when adding additional eigenvalues does not significantly increase this sum, the use of the corresponding eigenvectors as additional base vectors does not significantly improve the representation.

The optimal set of base vectors defined by this procedure is known in the statistical literature as the Karhunen-Loeve coordinate system.* The ADAPT processing of a collection of histories yields the components of the history vectors in this optimal base vector system, as well as the components of these base vectors themselves.

---

* For a complete description of the Karhunen-Loeve representation see Reference 1-A

For each history the NR components in the optimal system are the optimal representation of the data in the sense described above. Alternatively, the approach taken is conceptually analogous and numerical identical to finding a set of orthogonal functions to be used for a generalized Fourier series expansion of the original data histories. This problem is often encountered in classical boundary value problems of mathematical physics. In the case of the classical boundary value problem, the appropriate differential equation defines a set of orthonormal functions. To satisfy a given function on the boundary, this boundary function is expanded in the set of orthonormal functions which are defined by the differential equation. This set of orthonormal functions are optimal for representing this boundary condition in that they require less terms in the series to achieve a given degree of representation. This is equivalent to requiring less numbers to represent a given amount of information which is the exact criteria on which the Karhunen-Loeve expansion is based. In the case of empirical analysis, there is no differential equation to define the optimal set of orthogonal functions. However, the Karhunen-Loeve or eigen function expansion provides the numerical tool required to make any set of empirical learning data define its own set of optimal functions.

The optimal components derived in this manner are used in all further empirical analysis tasks performed by the ADAPT programs. Thus, the original M x N numbers representing M histories have been reduced to M x NR components, plus N x NR numbers to define the optimal vector base. Since the base system is optimal, the number of terms, NR, necessary to give a useful representation of a history is small, of the order of 10 to 30 and the reduction in the number of numbers is large, often of the order of 50 to 100.

In the process described so far, the optimal vectors are represented by their NC components in the first step orthogonal base, but this means they are a linear combination of the NC first step vectors, the coefficients being these NC components. Since these vectors are N-dimensional vectors, the optimal vectors can also be represented in the original N-dimensional space of the data history vectors by performing the linear combination.

The ADAPT representation process just outlined can be clarified with the simple example of two input histories, which is carried through analytically and described in appendices of References 5, 20, 21 and 32. For this special case the first optimal function is proportional to the average of the two history functions, the second to their difference, a result in accord with simple intuition. The relative sizes of the two eigen values is found to depend on the degree of correlation of the two histories, which has implications discussed later.

EVALUATION OF DATA USING ADAPT REPRESENTATION

Although the major objective of the representation is to reduce the dimensionality of the data for future processing it also provides an opportunity to better understand the nature of the data and to establish validity criteria for the application of the empirical analysis which may be performed using this learning data. The ADAPT programs provide the tools which are required to understand the quality of the base which has been derived. These tools are also useful in understanding the nature of the data. This understanding of the data provides a basis on which to select the dimensionality for the analysis. This section will present the tools which are available in the ADAPT programs for understanding the representation, analyzing the data and establishing empirical validity criteria.

A convenient measure of the degree of representation achieved with a given number of base vectors is the sum of the eigen values of the vectors used, divided by the average square magnitude of all of the original data history vectors. This represents the reduction in mean square error achieved divided by the total error reduction possible. In statistical terms this is the percent of the variation of the data explained by the representation used. If a set of data has zero variation, it does not contain any information. Extending this concept, we see that information must be at least monotonically related to the variation in the data. Furthermore, the variation has the form of an energy. Thus, the ratio of the sum of the eigen values divided by the average square magnitude of the original data histories is defined as the information energy. The ADAPT programs plot the percent information energy versus the number of dimensions used. These information energy curves are useful in at least three ways: 1) they provide a basis for evaluating the quality of representation which has been achieved, 2) they provide a basis to determine the different types of information which are available in the data set and 3) they provide a basis for selecting a dimensionality to use for the data analysis.

Fig 1-A presents the information energy curve obtained from a data set consisting of 50 histories of approximately 200 measurements each. These data histories were taken from Reference 34 and contain both dimensional measurements on diesel capsule valves and measurements of the performance of these capsule valves. There are two separate curves shown on this figure which have their initial point in common. The lower curve on this figure is the ratio of the eigen value associated with the optimal function indicated on the abcissa to the sum of all the eigen vectors or the sum of the square magnitudes of the learning data vectors. The upper curve is the cumulative sum of this ratio. This particular information energy curve is an example of a complete base in terms of the first step of the ADAPT optimization. That is for this case the first step of the ADAPT optimization is identical to that which would have been obtained using the classical Gram-Schmidt procedure. This is demonstrated by the fact that the cumulative information energy (given by the upper curve) reaches 100% of the available information.

The first point on Fig 1-A which is common to both the upper and lower curves shows that the first term in the optimal representation explains approximately 12% of the variation in the original data set. The third value on the lower curve shows that the third dimension explains approximately 6% of the variation in the data set and the corresponding point on the upper curve shows that the first three terms taken together explain approximately 27% of the information in the data set. Both the consideration of the characteristics of noise and the results of two case closed form solutions indicate that the lower or term by term information energy curve should be flat for random noise. This can be seen by noting that there should be no preferred optimal directions or functions for representing noise and each function should explain an equal amount of the variation when the variation is due to a random noise. Thus, in a complete base those terms lying above the point at which there is no further change in the slope of the lower curve of the information energy plot, are terms which are dominated by noise and should not be included in an analysis. For Fig 1-A this point occurs at a dimensionality of approximately 25. Examination of the upper curve corresponding to this dimensionality shows that only about 80% of the information in this data set is usable for analysis and the maximum useful dimensionality is 25. Note the ADAPT approach to determining the maximum usable dimensionality for the eigen value expansion is based on the quality of the information which is available, rather than the classical approach of simply assuming that when the eigen values have fallen below some small but arbitrarily selected percentage of the largest eigen value the analysis should be discontinued.

A second feature appearing in the lower curve on Fig 1-A is the break in the slope or knee occurring at the third dimension. This implies that the change of correlation between the information contained in the second and third terms of the representation is extremely great. This normally occurs when the phenomena being represented by those terms changes significantly. For this particular set of data, one can determine the phenomena causing this knee by examination of the projection of the entire set of learning data on the first two dimensions of the optimal space. This projection illustrates another analysis tool which the ADAPT programs provide. This tool is a plot of the projection of the entire learning space on to a plane defined by two of the optimal coordinates. Since the first two optimal coordinates together represent those two coordinates which contain the greatest amount of information possible for the learning data set, the projection of the learning data on to these two coordinate directions represent the best possible two dimensional display of all of the information contained in the learning data. Fig 2-A presents such a display for the data used to make the learning base from which Fig 1-A was obtained. Examination of the information energy plot shown in Figure 1 indicates that Fig 2-A represents approximately 20% of the information contained in the entire learning set. Fig 2-A may also be interpreted as a scatter plot of the coefficients of the first and second terms in the generalized Fourier series representation of each of the data histories using the optimal empirical orthogonal functions which are identical to the eigen functions derived in the Karhunen-Loeve procedure. Based on this interpretation, it follows that one can form the two term reconstruction of any of the learning data histories simply by multiplying the first optimal function by the coefficient along the NP1 direction and adding this to the product of the second optimal function times the coefficient along the NP2 direction.

Examination of Fig 2-A shows that the variation in these two dimensions is dominated by differences in the models of the engines in which the capsule valves were used. Examination of similar scatter plots for high order dimensions shows that this natural grouping according to model no longer occurs at the higher dimensions. Thus, the knee occurring at third optimal function in the information energy curve for this data set is due to the fact that the variation introduced by the different models of engines is contained in the first two terms of the optimal representation. This illustrates that the information energy curve can be used to select candidate dimensionalities for different types of problems. For example, if one were planning to use this data to study the differences between the model of engine used then two dimensions would probably be adequate for the analysis. However, if the data were to be used for the analysis of some other features associated with this data, the first two dimensions probably would not contain all of the pertinent information and a dimensionality between 2 and 25 would be required.

Fig 3-A presents an information energy curve for a base which although complete in terms of all of the dominant information does not fully represent all of the uncorrelated or noise-like information. This base is as useful for an analysis as the complete base which is illustrated by the information energy curve presented in Fig 1-A The information which has been omitted by this base due to the first step in the ADAPT selection criteria is only that information which is of a noise-like character. This can be seen by noting that the slope in information energy curve has become essentially zero after approximately the 40 to 45th term in the series.

Fig 4-A presents the information energy curve for a poor base. This curve shows an approximately constant rate of change of curvature over the majority of the higher dimensional portion of this curve. This suggest that one is still observing a change in the degree of correlation of the information as one increases the dimensionality. This implies that there is still useful information being added by increasing the dimensionality. However, the dimensionality can not be increased beyond that which was admitted by the first step selection criteria. This suggests that better results could be obtained if the base were re-derived using a different selection criteria for the first step in the ADAPT analysis. Even for a relatively poor base as illustrated by this information energy curve, the leading terms will be essentially identical to those that will be obtained with the complete first step base. Thus, the dominant results which can be obtained from the earlier terms in the series are valid. The results obtained using relatively high dimensionality with a base such as this, will generally be inferior to those obtained using a complete base.

The ADAPT empirical validity criteria is based on a measure of representation which may be applied to individual data vectors. This measure is the ratio of the square magnitude of the particular data in the optimal base to its square magnitude in the original data base. This is a measure of the information content which is lost as one transforms the data vector from the original space to the new space. This quantity is provided for each of the learning cases as a part of the standard ADAPT analysis. This test may also be applied to any test case to which one intends to apply the empirical algorithms which have been derived using this ADAPT representation. If this ratio is significantly smaller for the test cases then it was for the learning case there is a significant difference between the test data and the learning data and one is not justified in applying the empirical analysis to that test data. Thus, this ratio when applied to test histories serves as a basis for an apriori test of the validity of performing the empirical data analysis on that test case.

Fig 1-A - EXAMPLE OF AN INFORMATION ENERGY CURVE FOR A COMPLETE BASE

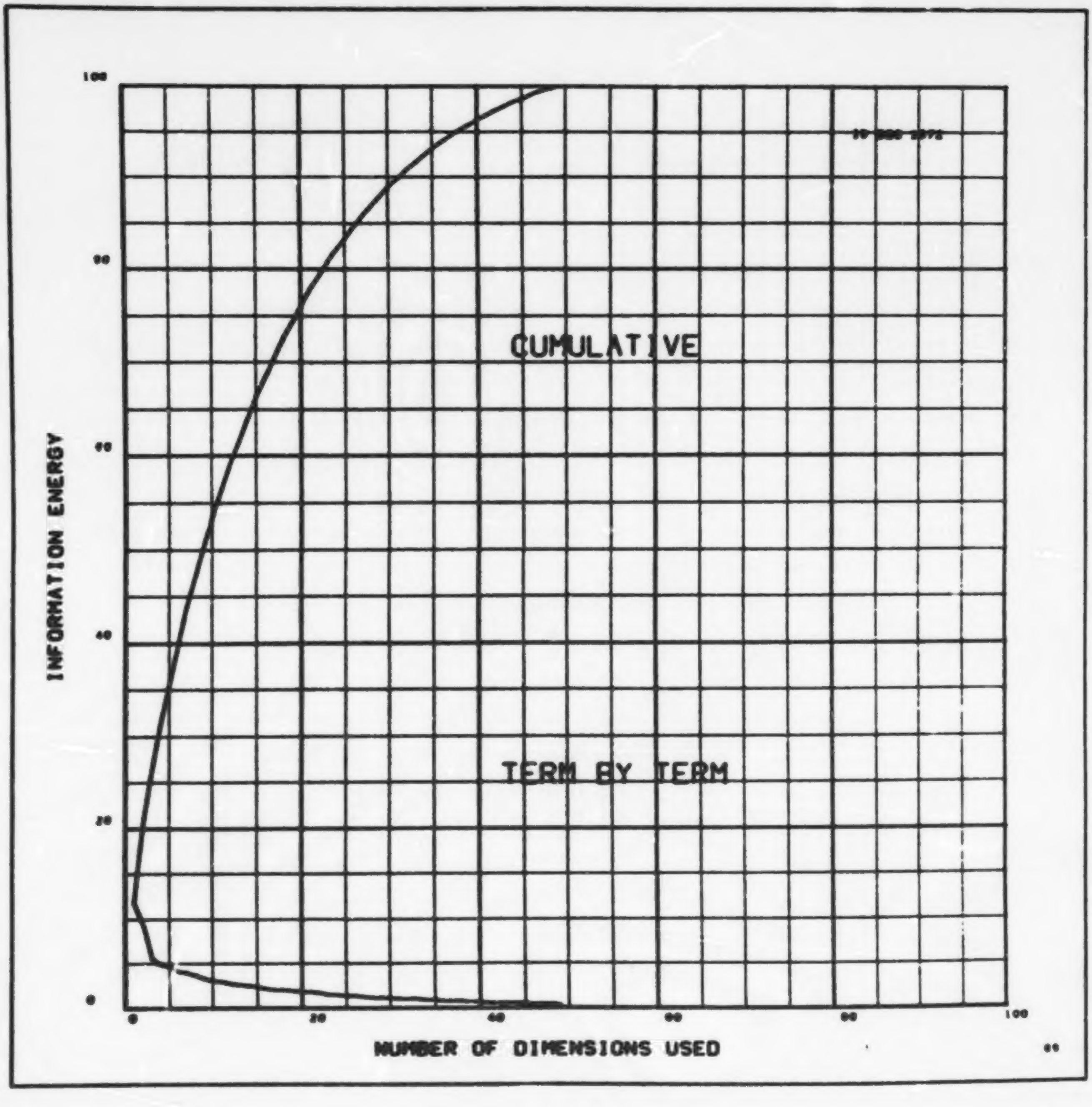


Fig 2-A - EXAMPLE OF A SCATTER PLOT OF THE FIRST AND SECOND OPTIMAL DIRECTIONS OF A BASE FOR REPRESENTING DIESEL CAPSULE VALVE DATA

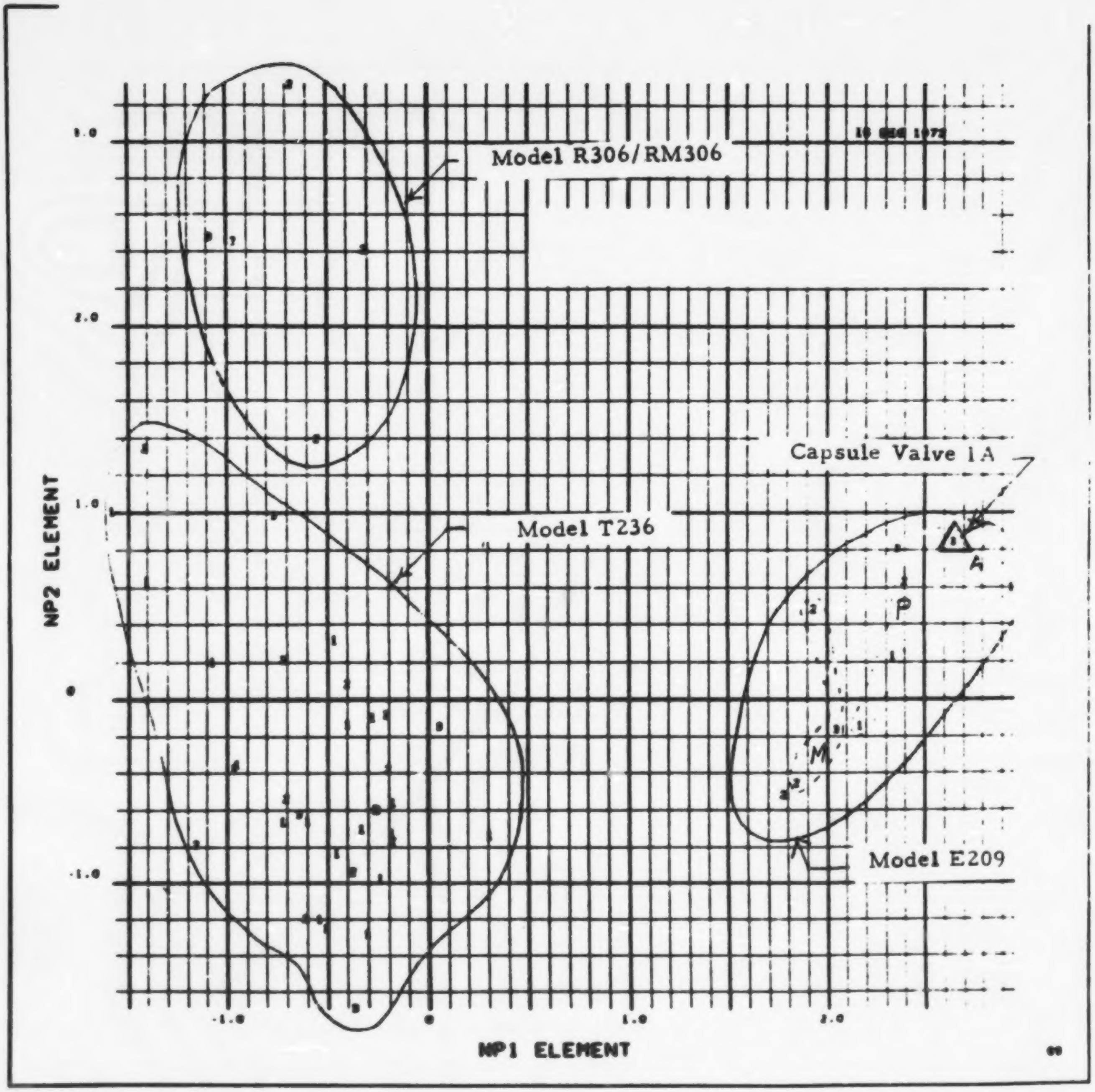


Fig 3-A - EXAMPLE OF INFORMATION ENERGY CURVE
FOR A GOOD INCOMPLETE BASE

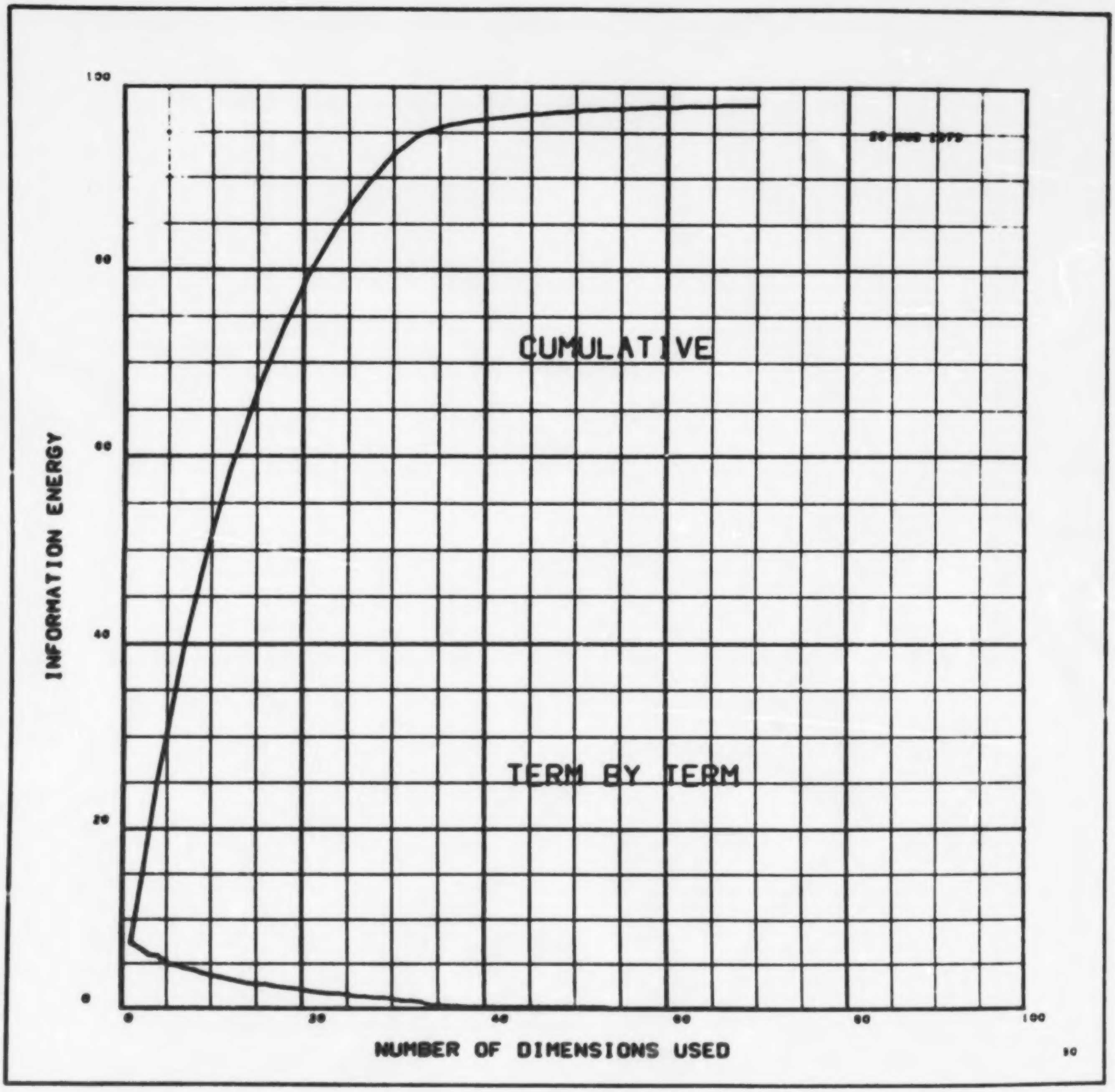
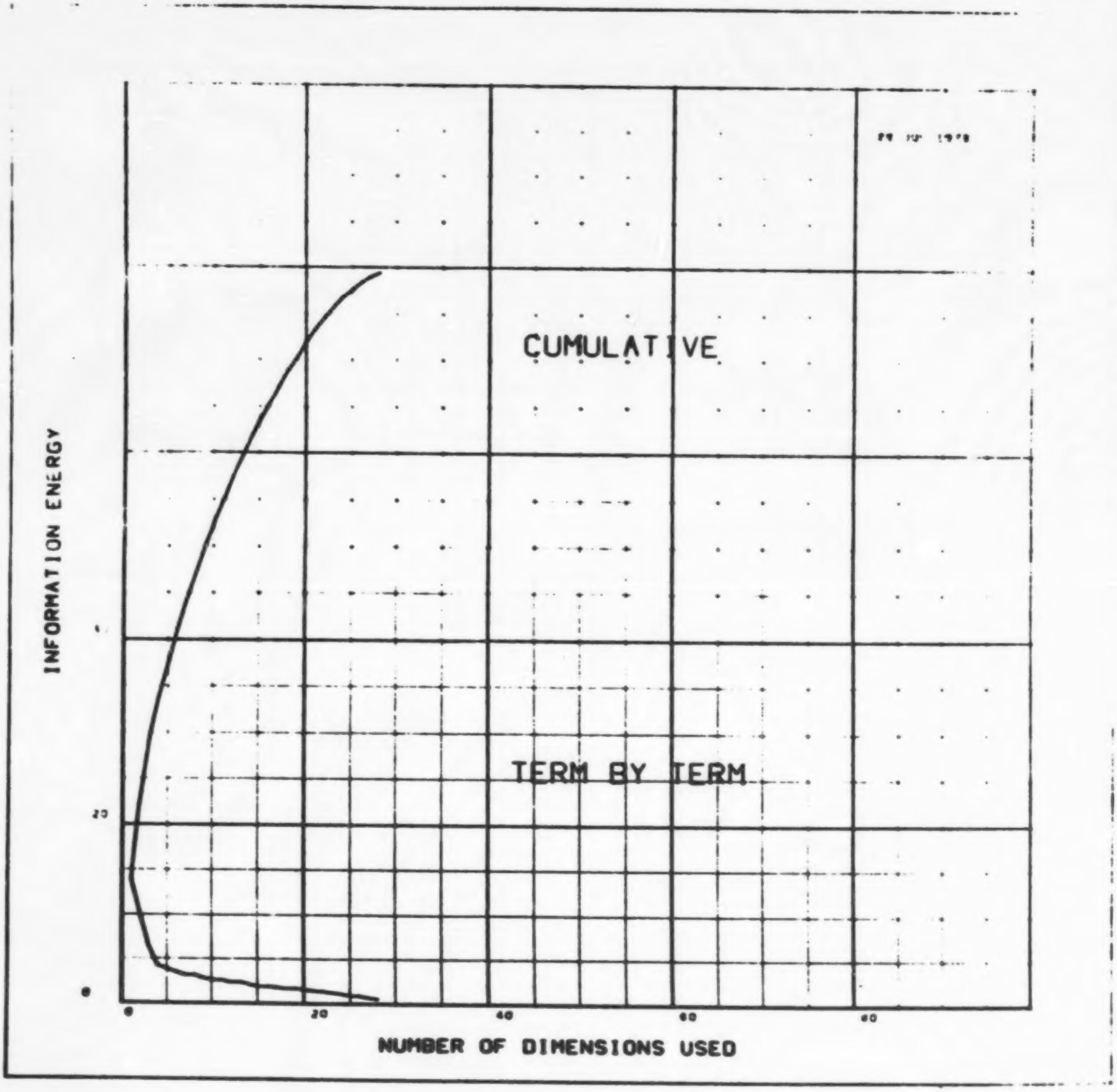


Fig 4-A - EXAMPLE OF AN INFORMATION ENERGY CURVE FOR A POOR INCOMPLETE BASE



APPLICATIONS USING OPTIMAL REPRESENTATION

Having arrived at the optimal (in the Karhunen-Loeve sense) representation, attention may now be turned to the use of the components of each of the data histories in this optimal space to accomplish the desired empirical analysis. Two of the more common forms of empirical analysis to which the ADAPT programs have been applied are classification or pattern recognition and parameter estimation or regression. The optimal representation not only greatly simplifies the classical approach to this empirical analysis but also provides additional capabilities. The use of ADAPT to accomplish these types of analysis will be discussed in the following two sections. The optimal representation is also extremely useful for other empirical analysis including such tasks as clustering, modeling and extrapolation. The ADAPT representation provides unique opportunities for empirical clutter subtraction and compacting of data.

CLASSIFICATION ANALYSIS

The derivation of classification algorithms in the optimal space is benefited by the fact that both the coordinates of the individual cases being studied and the statistics associated with these cases are expressed by fewer numbers. The covariance matrix which defines the statistics of the ensemble of the learning data is now a matrix having a dimensionality of NR by NR rather than the original dimensionality of N by N. Thus, the amount of analysis involving the covariance matrix is reduced by the square of the reduction in the dimensionality. Furthermore, the orthogonal properties of the space and the knowledge of the eigen functions can greatly simplify the derivation of some of the classical discriminates.

The simplest derivation of classification schemes incorporated in the ADAPT programs is simply visual examination of the scatter plots such as that shown in Fig 2-A. Had the objective of that analysis been to identify the engine from which the capsule valve was taken, the classification law could easily have been specified by examination of this scatter plot. This approach to the derivation of the classification laws is perhaps the most effective so long as one can deal with two or possibly three dimensions. However, for higher dimensionalities the inability to visualize separation surfaces requires the use of more formal approaches. The approaches which have been used for ADAPT analyses and are part of ADAPT include Fisher classification schemes, special classification schemes for populations having equal means, maximum likelyhood, Eckart and energy detectors. Other linear and non linear schemes may also be incorporated as required and will benefit by the simplifications noted above. Since space does not allow detailed descrption of all of the classification schemes which have been applied with the ADAPT programs the remainder of this section will be confined to the descriptions of the Fisher discriminant as it is incorporated in the ADAPT programs. Since many of the features discussed are common to any classification scheme, it will serve to illustrate the tools which can be made available with the ADAPT approach. The Fisher discriminant has been selected because experiences show that for many problems the Fisher discriminant is the most effective. An extensive comparison of the Fisher discriminant with other linear and non-linear discriminants has been performed and is presented in Reference 30.

The Fisher classification scheme is a linear classifier which seeks a line on which to project all of the data. This line is selected to satisfy the criteria that when all of the learning data histories are projected on this line the ratio of the distance between the means of the two classes divided by the sum of the dispersions of each of the classes is maximized. In the original derivation of this classification scheme by Fisher, the dispersions of each of the classes was equally weighted based on the number of members of that class. However, there are conditions under which it is advisable to use different weightings for each of these classes. It is this more general scheme which is included in the ADAPT programs.

SETTING OF THRESHOLD

The approach to setting the threshold to be used to classify the projection value obtained from applying the Fisher discriminant is based on the analysis presented by Anderson and Bahadur in

Reference 3. Strictly speaking, this analysis requires that all possible projection vectors produce Gaussian projections. In general, this is only true if the input data is itself Gaussian. For the great majority of projection directions, in particular those directions which are normally determined by the application of the Fisher discriminant, the Central Limit Theorem will result in a Gaussian projection. Thus, although the theory is not rigorously applicable, it is usually applicable to a large percentage of the possible projection directions when the data space is sufficiently large to invoke the Central Limit Theorem. Thus, one suspects it may still be a valid guide as to the selection of the Fisher weighting parameter and the threshold to be used with the Fisher discriminant. Experience with a great variety of data has shown that this is indeed the case.

Reference 3 shows that if one desires to minimize total number of errors made by the Fisher classification algorithm, one should select the Fisher weighting parameter, P, according to the following relation:

$$P\sigma_1 = (1 - P)\,\sigma_2 \qquad (2)$$

where σ_1 and σ_2 are the standard deviation of the projection values of the first and second classes, respectively. Assuming that the origin has been selected mid-way between the means of the projection values of each class the threshold, TH, is given by:

$$TH = (\tfrac{1}{2} - P)\ V \qquad (3)$$

Another criteria which one may wish to use, rather than minimizing the number of errors, establishes an algorithm which will achieve a desired false alarm rate. This special case is also discussed in Reference 3. Suppose one desires a probability $\bar{P}_N$, that there will be no false alarms in Class 1 when N Class 1 cases are examined (i.e. no Class 1 cases will be classified as belonging to Class 2). The following relation will define the false alarm probability for Class 1, P^I_{FA}:

$$(1 - P^I_{FA})^N = \bar{P}_N \qquad (4)$$

Solving this equation for the probability of false alarm for Class 1 under the assumption that $\bar{P}_N$ = is equal to 0.5 gives:

$$P^I_{FA} = 1 - \exp\,(\ln \bar{P}_N/N)\ \underset{N\uparrow\infty}{\sim}\ \frac{0.693}{N} \qquad (5)$$

Once the desired false alarm rate has been defined, Reference 3 shows that the proper Fisher weighting parameter to achieve this false alarm rate is given by:

$$P\,\sigma_1 = \beta^I \,(1 - P_{FA}^{I}) \qquad (6)$$

where β^I is the variable in the cumulative standard normal distribution function of the probability $1 - P_{FA}$. The corresponding threshold is given by:

$$TH = \mu_1 - \beta^I \sigma_1 \qquad (7)$$

where μ_1 is the mean of Class 1 and σ_1 is the standard deviation of Class 1.

The above equations, although strictly valid only for the case of Gaussian data, may be expected to give a good approximation even in the case where the data is not Gaussian, when the data space is relatively large. Experience with the utilization of these equations in a large number of real problems has verified that they do provide a good guidance for the selection of both the Fisher weighting parameter and the best threshold to achieve either the goal of minimum errors or a predefined false alarm rate.

Analysis of Classification Law

The procedure for deriving the Fisher discriminant in the ADAPT programs consists of the following steps: 1) the use of the learning data to derive the optimal representation, 2) the projection of all of the learning data into the optimal space, 3) the use of the learning cases represented in the optimal space to derive the Fisher classification direction and 4) the transformation of the Fisher classification direction back to the original data space. The first step has already been discussed in detail in Section 3. The projection of any learning case into the optimal space is accomplished by taking a dot product of the learning data vector with each of the base vectors of the optimal space.

The derivation of the Fisher discriminant using this representation yields a direction or a line in the optimal space. The components of this line may be considered as a spectrum indicating the importance of each of the optimal directions to the classification which will be performed. The square of each of these components have been plotted in Fig 5-A for an algorithm derived for the separation of diesel capsule valves which could be expected to have high fuel flow rates from those which could be expected to have low fuel flow rates. This figure shows that the most important dimensions for performing this separation was the 10th, 12th, 13th, 14th and 20th. This provides information regarding the effect of reducing dimensionality on the availability of the pertinent information for the decision. It may be used in conjunction with the information energy curve discussed in Section 3 to reach decisions as to the dimensionality which should be used for the analysis. However, since this line is invariant under coordinate transformation its coefficients in the original data space may also be obtained by transforming the line back to the original data space. The transformation between the original and the optimal data space is defined by the ADAPT optimal representation. Since this transformation is defined by an orthogonal matrix, the inverse of this transformation which is required to go from the optimal space to the original data space is the transpose of the matrix of the optimal functions. Thus, one may simply transform the Fisher classification line from the optimal space to the original space and examine the importance of each of the original measurements to the decision.

A plot of the line in Fig 5-A when transformed back in the original data space is presented in Fig 6-A For the problem each indexing variable may be associated with a specific measurement or tolerance in the capsule valve. For this study these measurements were grouped as indicated by the labels: fuel pressure, spring (SP), Seat (ST), etc. This plot has been given the name of the relative importance vector since it defines the importance of each of the independent variables to the decision which is made using the algorithm. It also provides the capability to apply the algorithm to test cases in the original data space without first transforming the test cases to the optimal space. Since the algorithm is just the dot product of the relative importance vector with the data history the absolute magnitude of the value plotted on the relative importance vector is a measure of significance of a given variable to the decision. For example, if a given variable shown in Fig 6-A has value of zero relative importance, the value of this variable in any data history will be multiplied by zero when it is added into the detection statistic. Thus, this variable can have no influence on the decision. On the other hand, if the relative importance vector corresponding to a given variable has a very large negative or positive value, even a relatively small change in the corresponding variable in the data vector may have a significant effect on the detection statistic and therefore the decision which is reached. Thus, even a casual examination of Fig 6-A shows that most important factors controlling the fuel flow from the capsule valve are variables controlling the fuel pressure and dimensions of the valve seat.

Although the relative importance spectra and vectors illustrated in this section have been derived for the Fisher discriminant it is clear that they may be derived for any linear classification scheme. These outputs are standard outputs for all the linear classification schemes which are incorporated in the ADAPT programs and are used both to understand the role of the optimal coordinates and each of the original measurements. This often provides a basis for physically understanding how the algorithm works.

REGRESSION ANALYSIS

If one wishes to associate a data history with a number rather than a class, one classical approach is a multiple regression analysis. The ADAPT programs include both least square and canonical regression schemes which may be used to derive parameter estimation algorithms in the optimal space. Both of these schemes require the inversion of matrices whose dimensionality is the square of the dimensions of the space in which the data is represented. Thus, once again the transformation of the data from the original space to the optimal space has resulted in a reduction in computation of the order of the square of the ratio of the dimensionalities of the optimal space to the original space. Since this ratio is often of the order of one to two orders of magnitude, we have again reduced the complexity of the derivation by several orders of magnitude. In many cases, this represents the difference between a feasible and an infeasible task.

The availability of the canonical regression scheme allows one to simultaneously fit any given data history to a number of dependent variables. However, in either case the algorithm derived is the dot product of the regression line with the data vector. Thus, as in the case of the linear classification laws, this line may be transformed from the optimal space to the original space and the algorithm may be applied in the original space without the necessity of transforming the data histories to the optimal space.

One is also able to form the relative importance spectrum, i.e. the components of the regression line in the optimal space, and a relative importance vector or the components of the regression line in the original data space. Fig 7-A presents an example of such a relative importance vector for a regression analysis for predicting the central pressure of a cyclone using the longitude of the storm, the latitude of the storm, the day of the year and 79 satellite observed temperature measurements. Although this display is very useful when dealing with a linear data history such as the diesel capsule valve problem, it is difficult to interpret a pictorial data history such as the effect of the temperature distribution on the estimate of the central pressure from this format. One can take the same transformation which was used to transform the picture to a linear data history and transform the relative importance vector back to the pictorial display. When this is accomplished one obtains a relative importance vector such as that shown in Fig. 7-A which shows the importance of each of the grid points on the radiation map to the calculation of the central pressure. Reference 8 shows that an experienced meteorologists can use this picture to understand the mechanisms which ADAPT has selected to predict the central pressure of the cyclone.

OTHER EMPIRICAL ANALYSIS--CLUSTERING ANALYSIS

In addition to the classification and regression analysis schemes which are incorporated in the ADAPT program, the optimal representation offers opportunities for other empirical analyses. One such opportunity is for clustering analyses which is often included under the general scope of pattern recognition. In the present discussion, we separate classification from clustering analysis in that classification analysis as used here refers to the derivation of a law to separate two apriori classes. On the other hand, a clustering analysis examines a set of data to determine one of the natural classes or groupings of the data which occur. After such clustering has been identified, one may derive a classification law to separate these clusters and examine the relative importance vector to determine the reasons for the cluster. Thus, clustering analysis is often a useful tool for evaluating the general nature of a set of data. It can also be useful in sub-dividing the data into sets for analysis.

One of the most useful tools for clustering analysis is the ADAPT scatter plot. In fact when very strong clustering occurs in the scatter plot of the first few optimal dimensions and when these first few optimal dimensions contain a large portion of the information energy it is usually desirable to make separate bases to analyze each of the clusters which are formed. One then has a two step analysis. The first step of the analysis uses the first few dimensions of the universal base to establish what cluster a given test data history belongs to. If the data histories vary over time this clustering is equivalent to finding the time epoch which is most appropriate for the particular data history. This epoch analysis may be considered as one step beyond classical trendline analysis. Rather than simply using the trend to update failure criteria, time clusters found by examination of the data allows one to account for discontinuities and non-linearities in the time variation when updating failure criteria.

In addition to the use of scatter plots to find clusters, the ADAPT programs also incorporate a nearest neighbor clustering scheme. This scheme is based on an algorithm which identifies those cases which are closest to one another in a high dimensional space. The performance of this algorithm can best be visualized by considering a nearest neighbor plot such as that presented in Fig 9-A This is a plot of the 50 capsule values used in the study presented in Reference 30. The capsule value number appears on both the abcissa and the ordinant of the plot. The ordinant is the capsule valve which is closest in the optimal space to the capsule valve listed on the abcissa. Thus, capsule valve No. 11 is closest to capsule valve No. 1, capsule valve No. 17 is closest to capsule valve No. 2, capsule valve No. 8 is the closest capsule valve to No. 3, capsule valve No. 43 is the closest capsule valve to No. 4, etc. When this plot has been constructed, it is read beginning with the ordinant instead of the abcissa. As an illustration consider capsule valve 6 as the starting capsule valve. If we read starting at an ordinate value of 6 we find that capsule valve No. 6 is the closest capsule valve to capsule valve No. 8, 13, 18 and 49. This is shown in the second tree of the Fig 10-A where Capsule Valve 6 has attached to it Valves 8, 13, 18 and 49. If we then examine the ordinate corresponding to Capsule Valve 8 we see that Capsule Valve 8 is the closest capsule valve to Capsule Valves 3, 6, 7 and 27, forming the second branch of the tree. Examination of the ordinate for values of 13, 18 and 49 show that these capsule valves are not closest to any other capsule valves. Similarly, examination of the second branch in the tree shows only Capsule Valve 3 is closest to any other capsule valve and it is closest to Capsule Valve 44 which in turn is not the closest to any other capsule valve. Thus, we conclude that the members of this tree are Capsule Valves 6, 8, 13, 49, 3, 7, 27 and 44, form some natural cluster. It is clear that this process is not as effective in identifying clusters as the human eye, however, this process is applicable to any number of dimensions where as the human eye has serious limitations if one attempts to find clusters in more than three dimensions.

MODELING AND COMPACTING

The ADAPT representation provides the most efficient way to represent any given amount of the information in the data. Thus, if one is trying to decide on what features to use in constructing a model based on an observed data, the optimal representation contained in the ADAPT programs provide this answer. Similarly, the ADAPT representation provides a natural mechanism for compacting the amount of data which must be stored or transformed when there exist an adequate empirical base to derive this optimal representation.

EXTRAPOLATION

The ADAPT representation provides a basis for extrapolating data histories. If one has a learning set consisting of a relatively large number of complete data histories, these histories may be used to construct a data base. If one now receives additional data histories which are incomplete, ADAPT programs exist to make a least square estimate of the coefficients which best fit that portion of the incomplete data history which is available. When these coefficients have been estimated, they may then be combined with the optimal functions derived from the complete data histories from the learning data to reconstruct the entire data history. Thus, the entire backlog of learning data is incorporated in the optimal functions. The available portion of the data history to be extrapolated is used in the finding of the least square fit to the best coefficient. This procedure has been applied to the continuation of both velocity altitude histories and to the extrapolation of the sunspot cycles which is reported in Reference 31.

CLUTTER SUBTRACTION

Clutter subtraction makes use of a modification of the first step in the formation of the optimal representation to eliminate certain directions from consideration in the optimization. This can be accomplished if the directions to be eliminated can be characterized. Classical situations for which this occurs include the effect of ground clutter on radar signatures and the effect of self-noise on sonar signatures. In these cases, one can obtain considerable amounts of data from the no target environment and utilize this data to determine the characteristics or the regions of the space in which the features which produce the clutter occur. If these regions of the space are then eliminated in the first step of the ADAPT optimization procedure they will not be available for consideration when the Karhunen-Loeve expansion is derived and the optimal representation will not include these regions of the space. This procedure is useful if one wishes to visually display reconstructed data histories without the clutter or if one wishes to use data from a number of sensors each of which have different clutter. In the latter case, clutter subtraction may be used to subtract clutter from each of the sensors prior to the comparison of the results.

FIGURE 5A- EXAMPLE OF A RELATIVE IMPORTANCE SPECTRUM FOR A FISHER CLASSIFICATION LAW

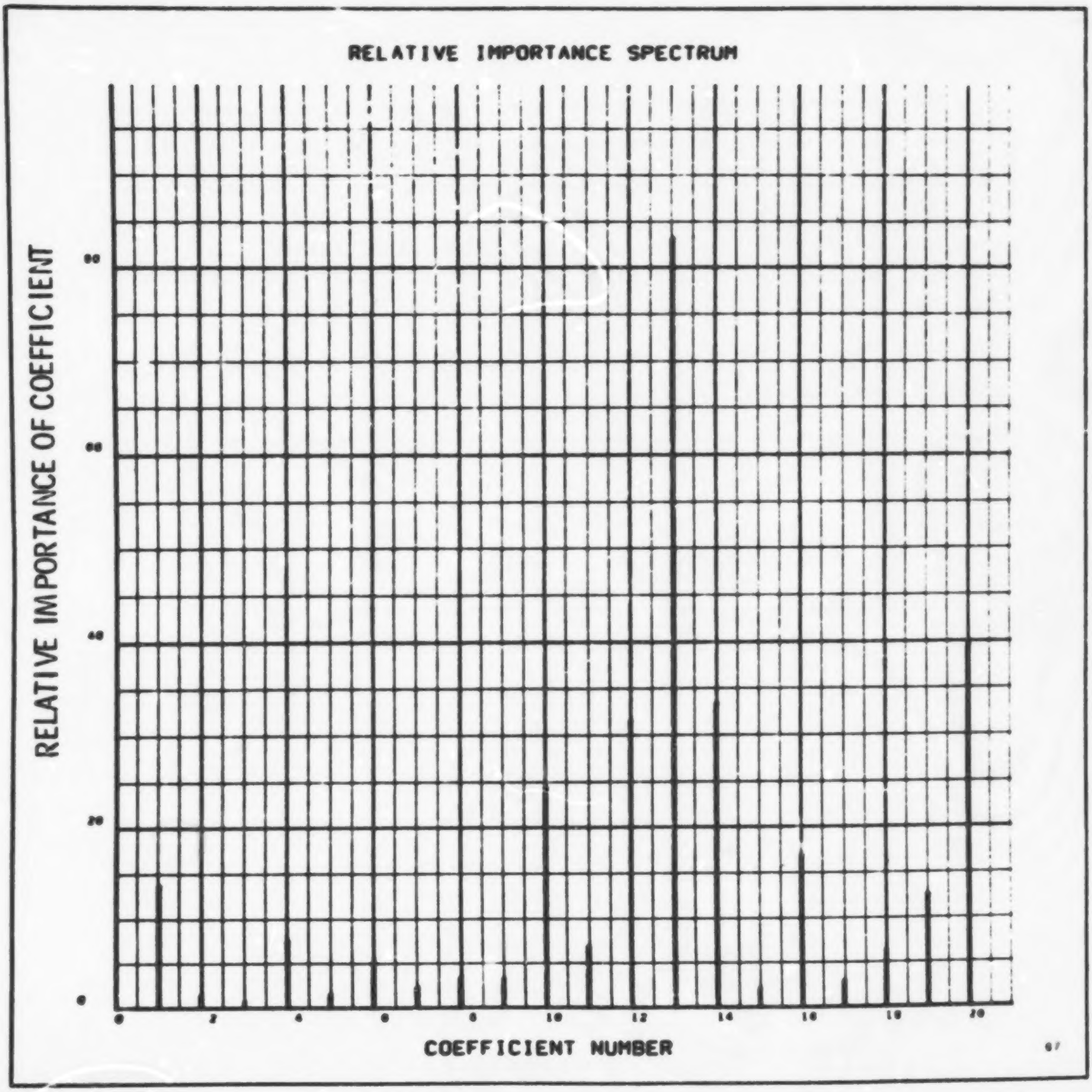


FIGURE 6A - EXAMPLE OF A RELATIVE IMPORTANCE VECTOR FOR A FISHER CLASSIFICATION LAW

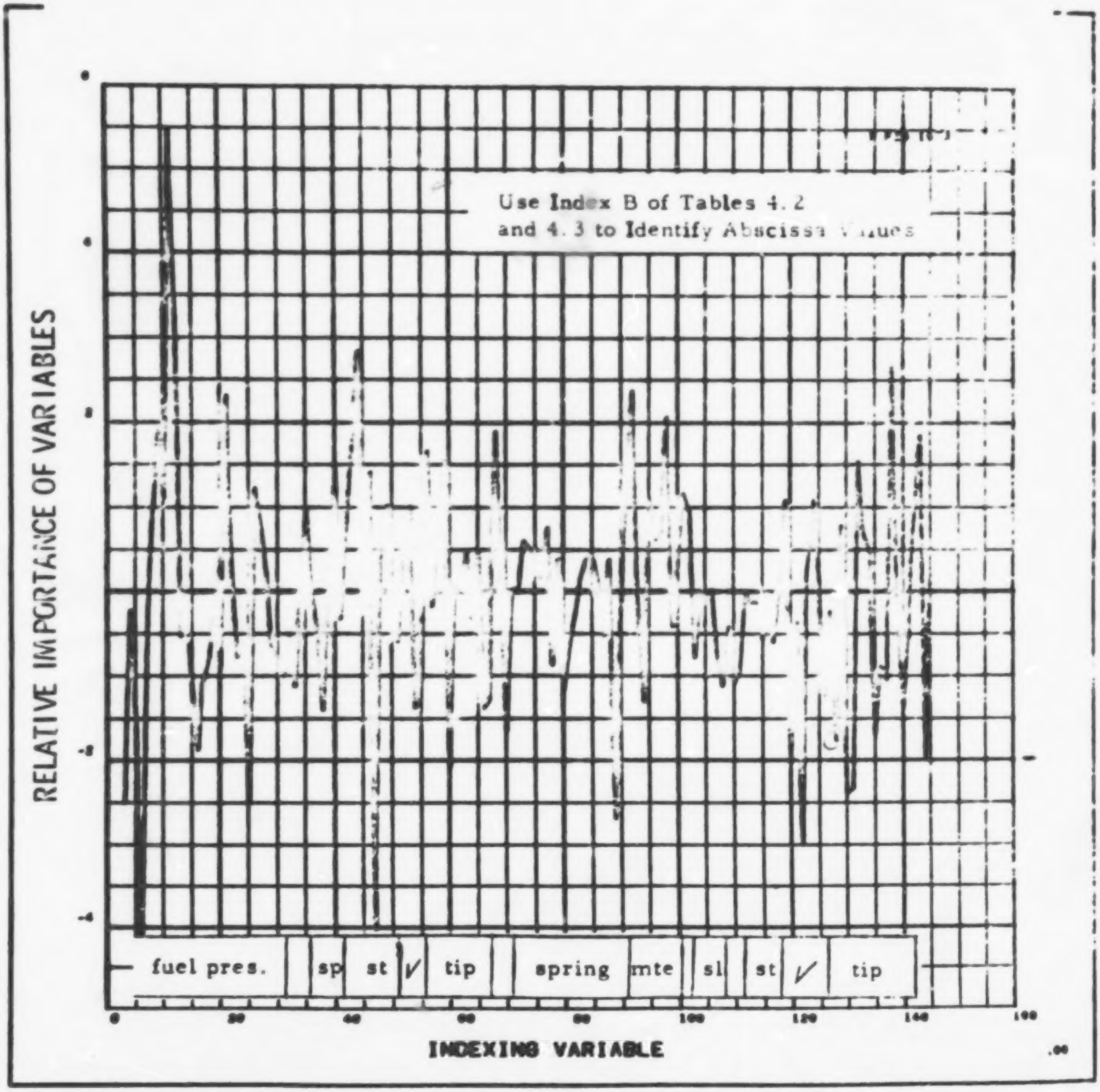


FIGURE 7A - EXAMPLE OF A RELATIVE IMPORTANCE VECTOR
FOR A REGRESSION LAW

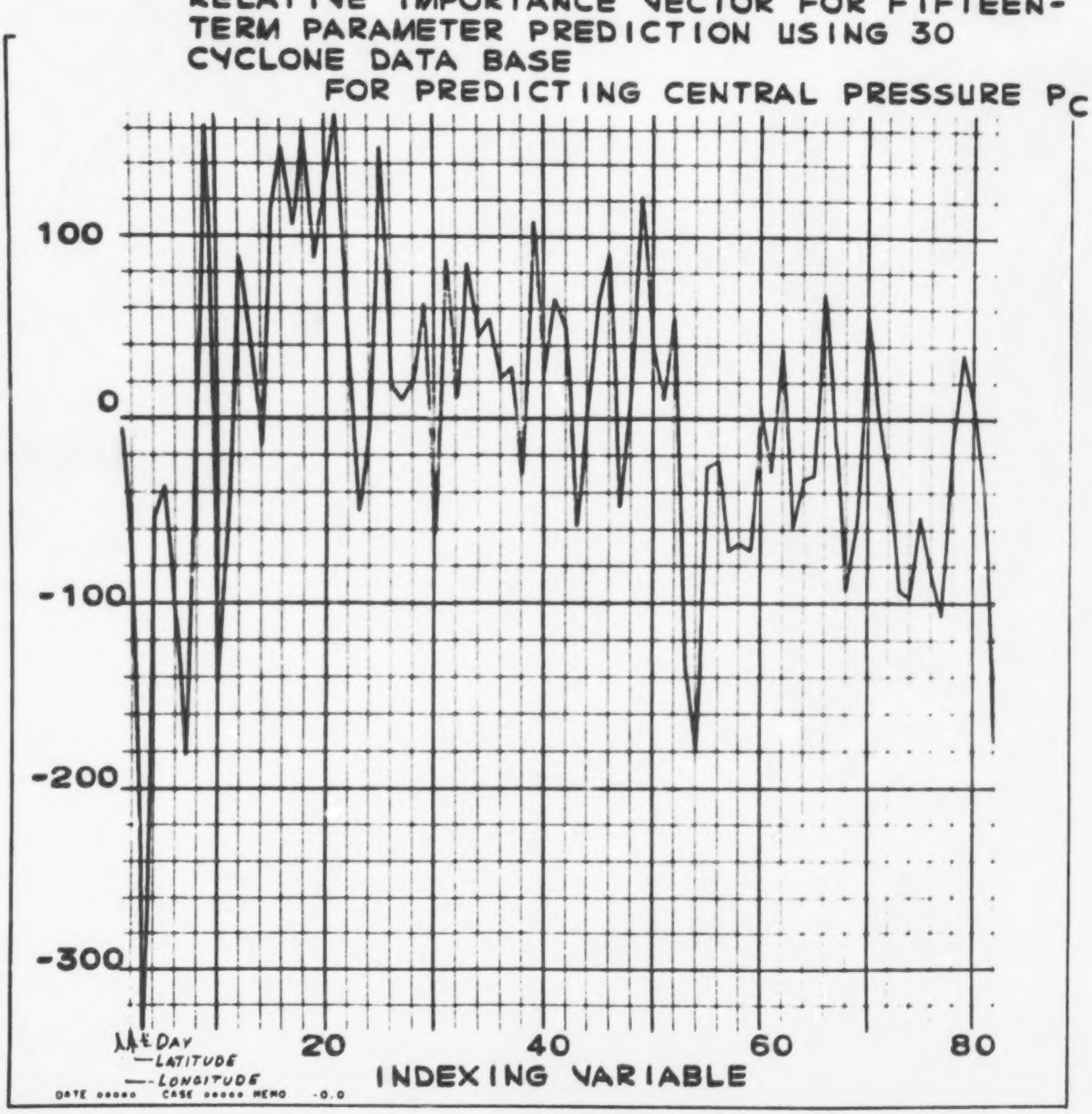
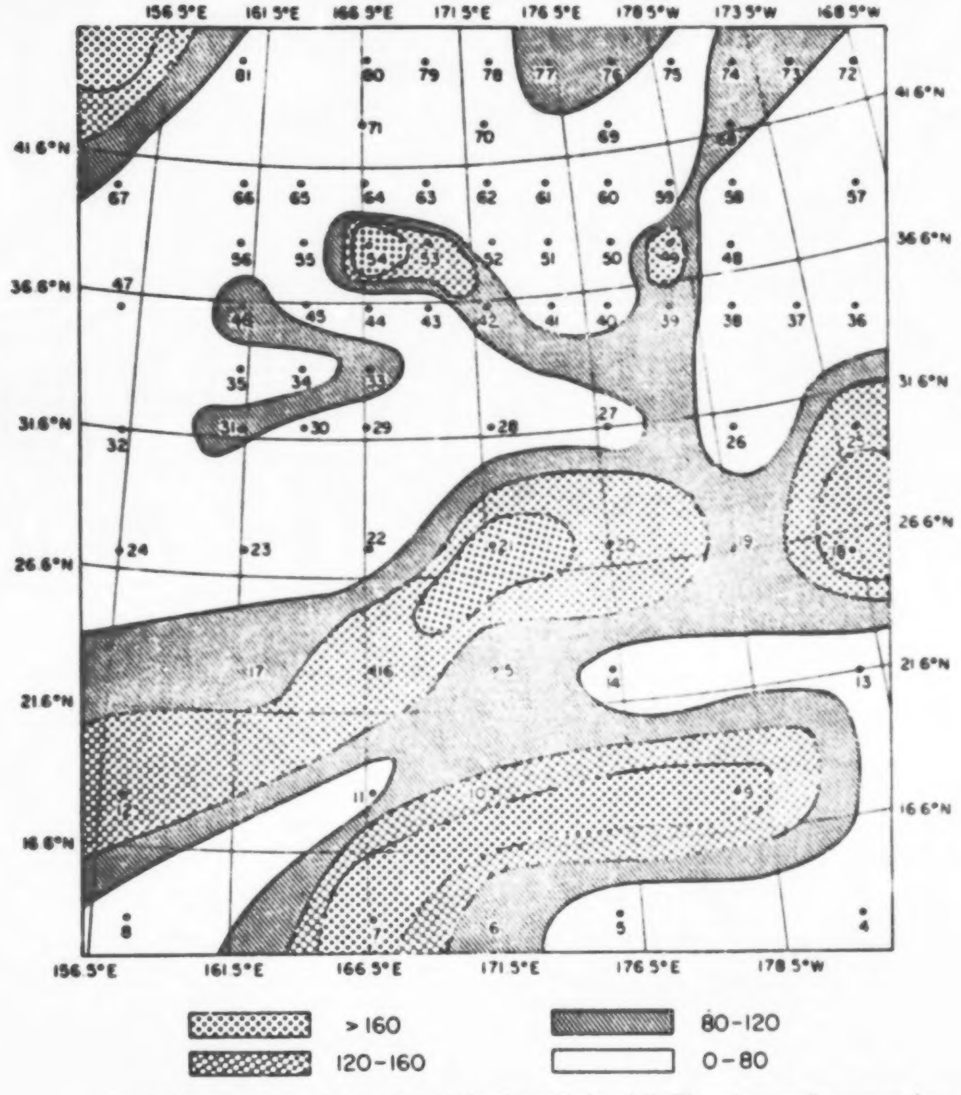


FIGURE 8A- EXAMPLE OF RELATIVE IMPORTANCE VECTOR
TRANSFORMED TO TWO DIMENSIONAL FORMAT


156.5°E
161.5°E
166.5°E
171.5°E
176.5°E
178.5°W
173.5°W
168.5°W
41.6°N
36.6°N
31.6°N
26.6°N
21.6°N
16.6°N
>160
120-160
80-120
0-80

The relative importance of each grid point for the prediction of P_e. The values are the same as those shown in Fig. 7 A grid point number is shown beneath each dot which identifies the grid point position.

FIGURE 9A - EXAMPLE OF A NEAREST NEIGHBOR PLOT USED TO CONSTRUCT A NEAREST NEIGHBOR TREE

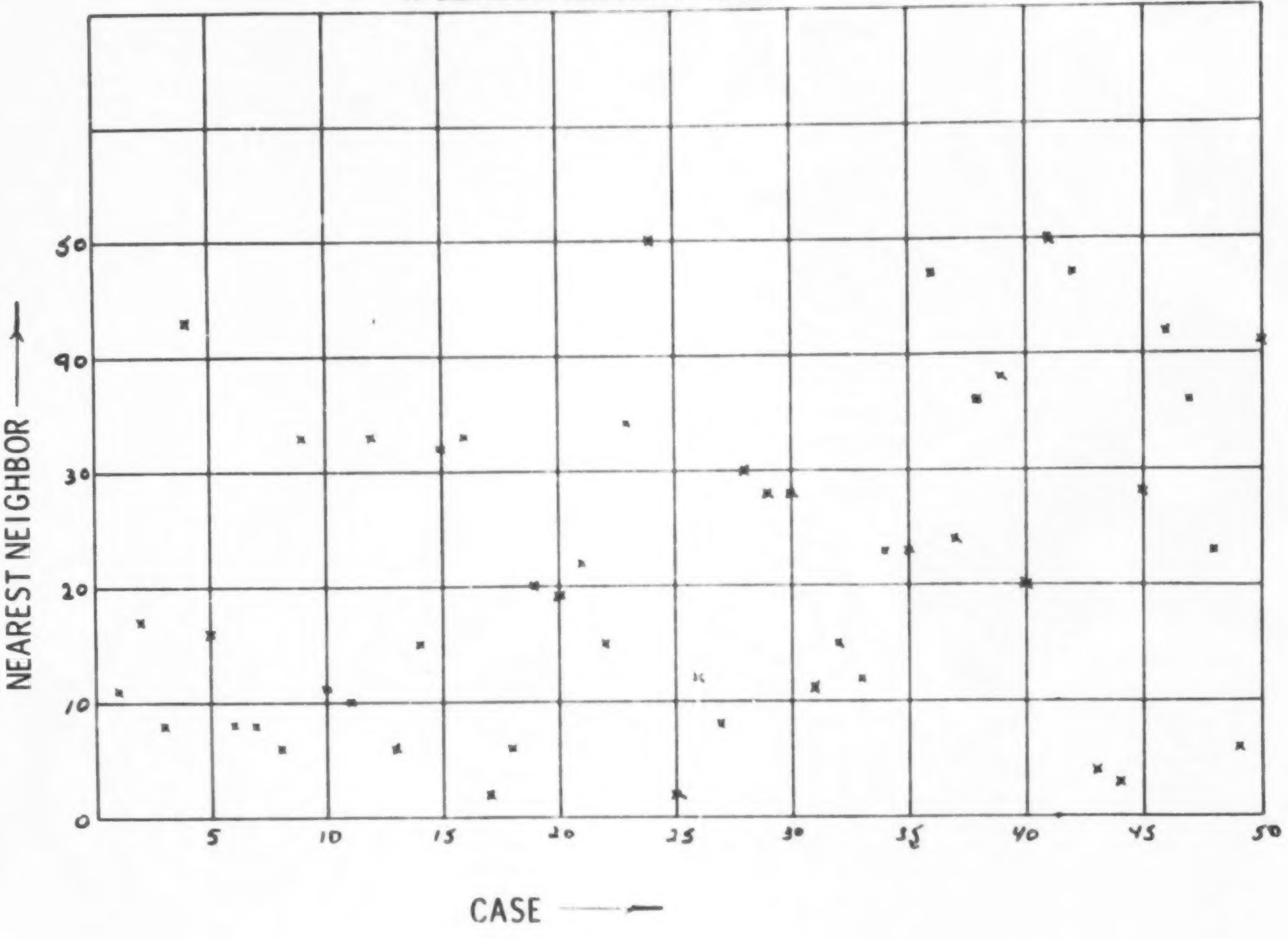
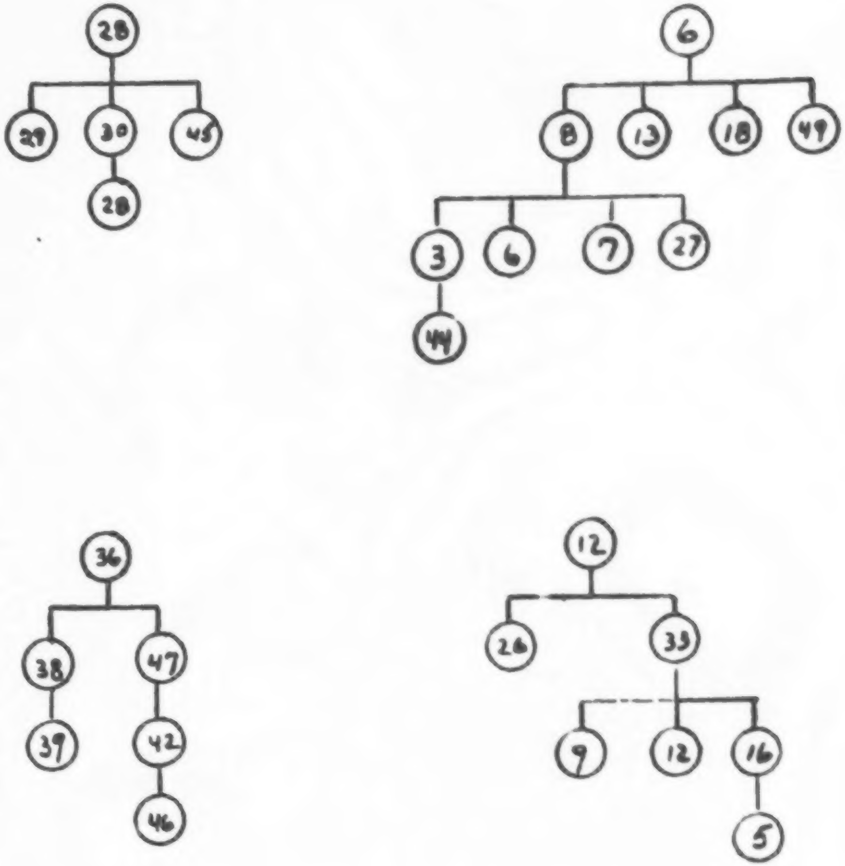


FIGURE 10A- EXAMPLE OF A NEAREST NEIGHBOR TREE


28
29
30
45
28
6
8
13
18
49
3
6
7
27
44
36
38
47
39
42
46
12
26
33
9
12
16
5

PERFORMANCE EVALUATION

The classical method of evaluating the performance of an empirical algorithm is to apply this algorithm to independent test cases and derive the performance statistics from the results of these independent tests. Although this is the only acceptable approach for demonstrating this performance to the scientific community it has several significant disadvantages as an analysis tool. The most important disadvantage is the cost and time required to perform the independent tests at each stage of the analysis. Thus, experience with empirical analysis has shown a need for developing a capability to estimate the performance that a given algorithm will achieve without the need for performing independent tests.

The ADAPT programs include procedures which allow the estimate of this performance based on the performance of the algorithm on the learning data. These procedures have been developed for both classification analysis and regression analysis. The procedures are based on the concept that when the ratio of number of learning cases to number of dimensions is sufficient the performance of the algorithm on the independent test cases will approach the performance on the learning data. Experience has shown that acceptable ratios of the number of cases to the number of dimensions used are dependent upon the performance of the algorithm. Thus, an experimental performance map has been developed which provides the analyst with a basis for estimating the performance of an empirical algorithm on independent test data from the performance of that algorithm on the learning data. This tool is utilized in the process of developing algorithms to reduce both the time and cost required to select the best approach to deriving the empirical algorithm. Final demonstration of the best algorithm which has been achieved is still accomplished by independent test, if test data can be obtained.

EVALUATION OF CLASSIFICATION PERFORMANCE

The simplest approach to visualize the performance of a classification algorithm is to examine the values obtained by application of the algorithm. This may be accomplished by presenting these values as a bar chart of the detection statistic versus each of the cases examined. This presentation is included in the ADAPT programs and is very useful to visualize the detailed characteristics of the performance. However, it is often desirable to be able to compare the performance of a large number of algorithms. One such situation is the study of the trade-off between detection probability and false alarm rate. The classical approach for accomplishing this study is to present the data in the form of receiver operating curves. These curves are simply plots of a detection probability versus the false alarm rate and can be obtained by evaluating the algorithm performance as one varies the threshold.

It is also difficult to evaluate the effect of dimensionality on the performance of algorithms by studying large numbers of bar charts. Thus, it is desirable to introduce a single measure of algorithm performance which can be used to study the effect of dimensionality on the performance of the algorithm. One convenient measure is the probability of error. This measure has the advantages of a simple intuitive meaning and has a unique relationship to the receiver operating curve. Thus, it is desirable to express the performance for the classification law in terms of the probability of error. This can be accomplished for the Fisher discriminant by examination of a quantity V. Since the Fisher discriminant is the result of a maximization of V, which can be defined by:

$$V = (\mu_1 - \mu_2)^2 / \left(P\sigma_1^2 + (1-P)\sigma_2^2\right) \qquad (8)$$

it is clear that the maximum value of V is itself a good measure of the performance of the algorithm. The maximum value of V, over all possible projections, turn out to occur when the denominator of Equation 8 equal to the square root of the numerator, which means V becomes, geometrically, the distance between the means of the projection of the two classes on the Fisher direction. Thus, for the Fisher discriminant Equation 8 provides a relationship between the projection of the means of the two classes, the standard deviation of each class, and the Fisher weighting parameter.

It is interesting to consider the special case in which the standard deviation of each of the classes is equal. For this case,

$$\sigma^2 = V = |\mu_1 - \mu_2| \qquad (9)$$

and

$$(\sigma_1 + \sigma_2) / V = 2/\sqrt{V} = 2\sigma/V \qquad (10)$$

This parameter is used as a measure of the goodness of performance of the discriminant. Regardless of the relationship between the standard deviations of the two classes, the smaller $\sum \sigma / V$ (the larger V) the better the performance of the algorithm. In this case where the standard deviations of both classes are equal V may be related directly to the probability of error, PE.

For the case where the threshold is selected to minimize the number of errors, the situation is shown in Figure 11A. The threshold is set half way between the mean projections of the two classes, because the criterion requires that the errors for the two classes are the same. Then the probability of error is the shaded Area A_1 which is value of the cumulative normal distribution centered on μ_1, up to $\mu_1 - V/2$. If G is the standard cumulative normal distribution, this is

$$P_E = G \left\{ \frac{-V/2}{\sigma} \right\} = G \left\{ \frac{1}{2/\sqrt{V}} \right\} \tag{11}$$

P_E is the probability of making an error in either class, and

$$P_D = 1 - P_E \tag{12}$$

is the probability of correctly identifying a member of either class.

For the case where one wishes the maximum detection of Class 2 for a specified error probability P_{FA}^{I} of Class I, the threshold is set by Equation 7. Again, for equal standard deviations, the situation is quite simple. If we take the origin half way between the mean projections of the two classes then $\mu_1 = V/2$, and Equation 7 becomes

$$TH = V/2 - \beta^{I} \sigma \tag{13}$$

$$\text{or} \quad \beta^{I} = (V/2 - TH)/\sigma \tag{14}$$

This is the standard normal deviate at which

$$P_{FA}^{I} = G(\beta^{I}) \tag{15}$$

The detection probability P_D^{II} of Class 2 is the area under the normal curve centered on $\mu_2 = -V/2$ up to TH. The normal deviate for this curve at that point is

$$\beta^{II} = (TH + V/2) \tag{16}$$

and

$$P_D^{II} = G(\beta^{II}) \qquad (17)$$

But TH can be eliminated from (14) and (16) to give

$$\beta^{I} = \sqrt{V} - \beta^{II} = \frac{2}{2\sqrt{V}} - \beta^{II} \qquad (18)$$

Thus, for the case of equal standard deviations, the detection probability of Class II depends only on the false alarm probability of Class I, and the Fisher maximum through $\sum \sigma / V$. Fig 12-A presents these ROC curves for various values of the parameter $\sum \sigma / V$.

In addition to obtaining an understanding of the trade-off between detection probability and false alarm rate, it is important to have a measure of algorithm performance to evaluate the effect of dimensionality of the space in which the algorithm is derived. This is extremely important since the use of too large a dimensionality in the derivation of an algorithm will result in the algorithm being derived by fitting the learning data according to special characteristics of the particular learning sample, and not according to characteristics of the population sample. That is, the major basis for the separation will be the difference between the population and sample means, rather than the difference between the means of the two populations being classified. When this occurs the classification may be called "overdetermined". This phenomena is quite analogous to the fitting of a third order polynomial through a set of data. If a third order polynomial is fit to 3 data points, there is no reason to believe that a general law has been derived. However, if this same third order polynomial makes a reasonably good fit to 100 points, there is little doubt that these 100 points are related by some phenomena which is well expressed by a third order polynomial.

Thus, it is important to understand the capabilities of a Fisher discriminant to derive classification algorithms simply on the difference between sample and population means. Originally the ADAPT analysis team evaluated this by performing separations of odd cases versus even cases from both classes for each problem being considered. The performance of these separations were then compared with the performance of the classification algorithm derived between the desired classes. If the algorithm derived for separating the odd versus even gave a similar performance to the desired algorithm then one concluded that the algorithm was not based on physical characteristics but rather on the differences between the sample and the population means and was considered to be overdetermined. This experience can be summarized in a plot such as presented in Fig 13-A This figure plots the number of cases divided by the number of dimensions versus the performance measure. The cross hatched curve is an experimental curve separating valid from overdetermined separations. It is based on separations of odd from even (i.e. random separations) for a large variety of problems and data. The extrapolation of this curve for low values of the performance measure was accomplished by making a similar plot on a linear scale and noting that for a number of cases over number of dimensions of unity the performance measure should go to 0. It is interesting to compare the cross hatched curve of Fig 13-A with the results of a similar analysis presented in Reference 4 which indicated that when the number of cases to number of dimensions exceeded six, one could have confidence in the performance of the algorithm. Fig 13-A shows why this is the case. Remembering that for $\sigma_1 = \sigma_2$ we may relate $\Sigma \sigma/V$ to the probability of error we note that for a performance measure of 2 the probability of error is approximately one in three. Since a random process for selecting a class has a probability of error of one in two, it is clear that an algorithm whose performance measure is two or greater is not interesting. Thus, this curve shows that any algorithm of interest derived in a space such that the number of cases divided by the number of dimensions is greater than six lies to the left of the experimental curve.

When an algorithm is derived using the Fisher discriminant it may be placed at some point on Fig 13-A by noting the number of cases used in the learning data, the number of dimensions of the space in which the algorithm is derived, and the performance measure for that algorithm. All of these parameters are available in the ADAPT output for the deviation of the Fisher discriminant. If the algorithm falls to the right of the cross hatched region in this figure, one knows that it is overdetermined and is not a valid algorithm. If it falls near but to the left of the cross hatched area, one realizes that the performance of this algorithm on the learning data is significantly better than one can expect on the test data. Only if the algorithm falls to the left of and reasonably far away from this cross hatched area does one have an algorithm whose learning data performance is indicative of the performance which can be expected on a test case.

It is useful to visualize the path of a typical algorithm on this performance map in conjunction with the other ADAPT analysis tools. If one were to examine the projection of all of the learning data on to the first optimal coordinate direction, one could determine for any particular classification law a probability of error or $\sum \sigma / V$ for the algorithm consisting of the projection on the first eigen vector. This would be the performance of any linear classifier derived using only the first optimal direction. If the desired separation is based on information which dominates the variation of the data set one might expect that this classification procedure would yield a useful result. If not the classification procedure may be of no value at all. In either case, it can be located on a performance map. If we assume that the first eigen direction has no bearing on the classification which is desired, then this point would fall to the right of the cross hatched curve on Figure 13A. As one now increases the number of dimensions used and repeats the application of a Fisher discriminant at each dimensionality, the curve will continue some path to the right of the cross hatched region until sufficient number of dimensions is used that some of the data pertinent to the desired classification is incorporated into the analysis. At this point the classification will no longer have the character of a purely random classification and the curve can be expected to move to the left of the cross hatched curve. Analysis of the relative importance spectrum would show that the first significant dimension was also reached at this point. It is also likely that the information energy plot will show a knee at this point in the curve indicating that a different type of data having a more noise-like characteristic is now being considered.

As one continues to increase the dimensionality, the performance of the algorithm may be expected to increase for two reasons. The first is that additional hopefully pertinent information is being added. At any dimensionality where pertinent information is added one can expect that the track of the algorithm on the performance map will become more horizontal and that this dimension will correspond to a large value in the relative importance spectrum. A second reason for improvement in the performance is that as the number of dimensions gets sufficiently large the algorithm will approach an overdetermined situation. The dimensionality at which this occurs is a function of the number of learning cases used and the performance of the algorithm. As the ratio of number of cases to number of dimensions approaches unity the path of the algorithm will again intersect the cross hatched line on Figure 13A.

The preceding discussion suggests that the ADAPT scatter plot and relative importance spectrum will provide a good estimate of the shape of the track of an algorithm for higher values of the ratio of number of cases to the number of dimensions then were used for the derivation of the algorithm. By examination of this track, one can estimate whether the performance achieved on the learning data will also be achieved on the test data and whether the dimensionality used contains any information which is pertinent to the desired classification. If one must use a relatively small ratio of number of cases to number of dimensions to achieve a valid classification, on the performance map, more learning data is needed to derive a useful algorithm. When an algorithm is to the left of the cross hatched curve the data does contain information which is pertinent to the desired classification. However, if this can only be accomplished at a small ratio of the number of cases to number of dimensions, additional learning cases would allow one to increase this ratio and still use a sufficiently high dimensionality to obtain the information which is pertinent to the desired classification. For specific examples illustrating tracks of algorithms on the performance map, the reader is referred to the discussions of performance maps presented in References 30, 32 through 34 and 36 through 38.

THE PERFORMANCE EVALUATION OF REGRESSION ANALYSIS

The performance evaluation of regression analysis is similar to that of the Fisher classification analysis discussed in the preceding section. The pictorial presentation of performance of a single algorithm corresponding to the bar chart is now a plot of the actual parameter versus the estimated value of the parameter. These plots can be prepared showing either the learning or the test data or both the learning and test data on the same plot. Perfect agreement on this plot results in the data falling on the line having a 45 degree slope and passing through the origin. This plot has the same disadvantages for evaluating large numbers of algorithms and for studying the effect of dimensionality on performance that the bar charts have for the Fisher discriminant.

It is useful to examine the performance of a regression law on the learning cases and estimate what the performance would be on the test cases. The ADAPT programs include performance maps for accomplishing this for regression analysis which are analogous to the performance maps presented for the Fisher classification algorithms. For the case of the regression analysis, the performance maps are a plot of the ratio of the number of cases to the number of dimensions versus the performance of the algorithm. The major difference is that the performance of the algorithm is now measured by the quantity σrat. This quantity is defined as the ratio of the standard deviation of the estimate about the actual value divided by the standard deviation about the mean value and is given by:

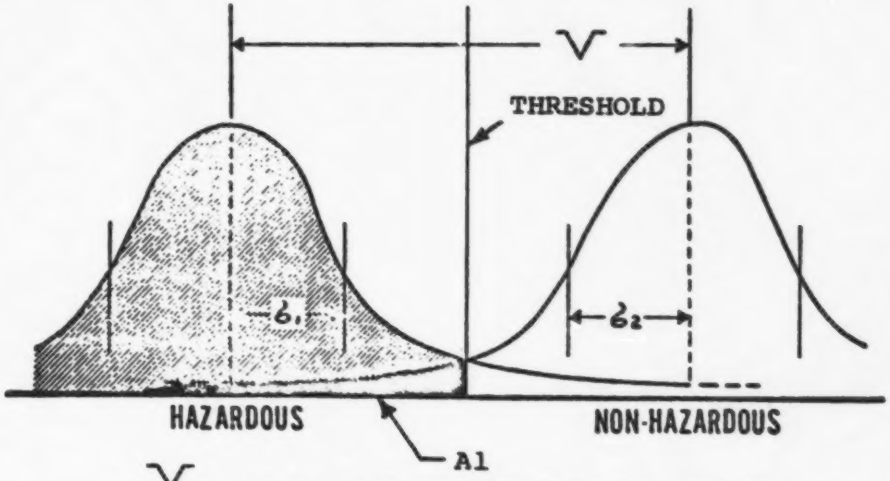
$$\sigma_{rat} = \frac{(V - Z)^2}{(V - \bar{V})} \qquad (19)$$

where:

V = actual value
Z = estimated value
$\bar{V}$ = mean of actual values

As in the case of the classification performance map, experimental analyses have been performed on a large variety of data which provide experimental curves as to the confidence that the algorithm derived is not overdetermined as a function of its location on a performance map. A performance map including these curves is presented in Fig 14-A. This regression performance map can be used in a manner exactly analgous to that explained for the classification performance map. The track of an algorithm on this performance map follows the same logic as was developed for the track of an algorithm on the classification performance map.

FIGURE - 11A
PERFORMANCE MEASURE FOR CLASSIFICATION



$\frac{V}{\sigma_1 + \sigma_2}$ = MEASURE OF QUALITY OF SEPARATION

[DETECTION PROBABILITY, P_D] = [AREA UNDER HAZARDOUS CURVE TO LEFT OF DETECTION LEVEL]

$$\left[\text{FALSE ALARM RATE}\right] = \frac{\left[\text{AREA UNDER NON-HAZARDOUS CURVE TO LEFT OF DETECTION LEVEL}\right]}{\left[\text{SUM OF AREAS UNDER HAZARDOUS AND NON-HAZARDOUS CURVES TO LEFT OF DETECTION LEVEL}\right]}$$

FIGURE - 12A

CLASSIFICATION PERFORMANCE TRADE-OFF CURVES FOR EQUAL-GAUSSIAN STANDARD DEVIATIONS

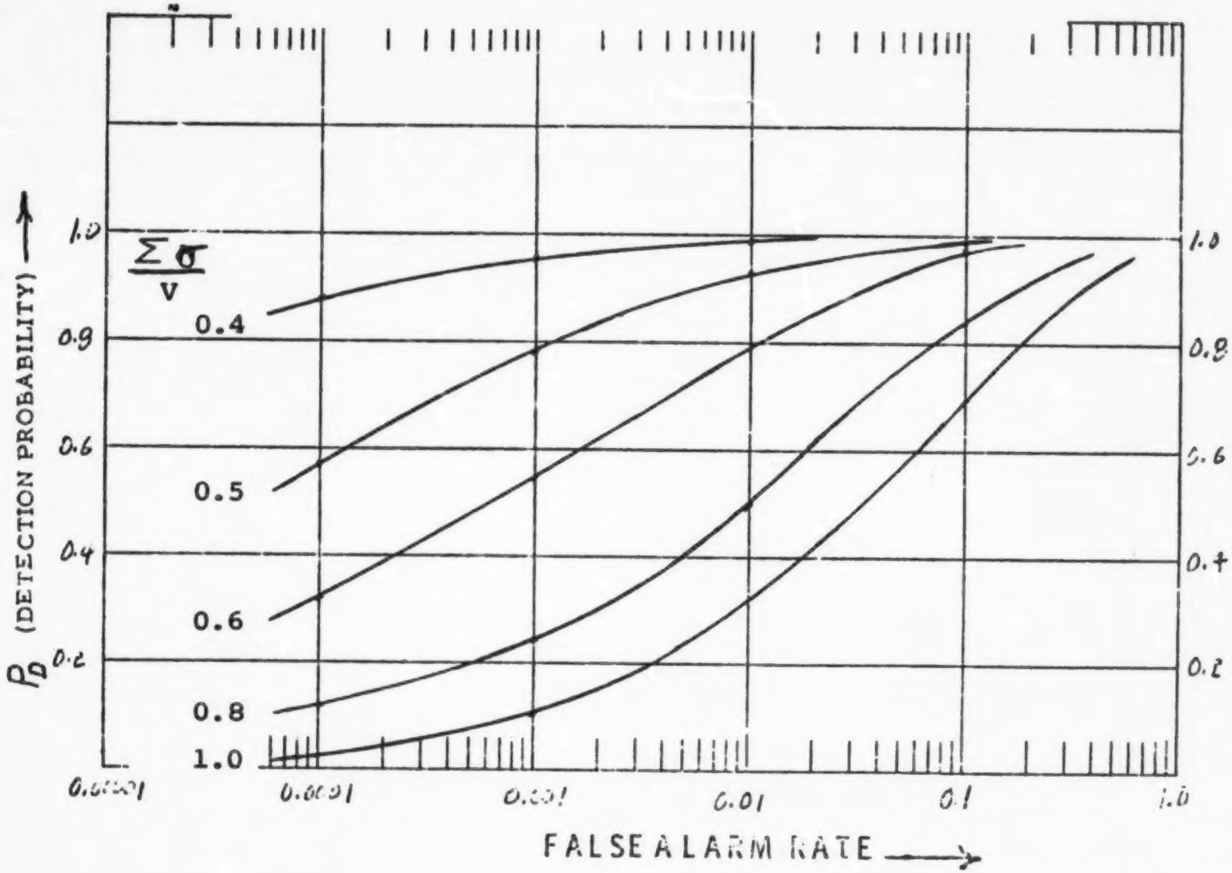


FIGURE- 13
PERFORMANCE MAP

FOR FISHER CLASSIFICATION ALGORITHMS

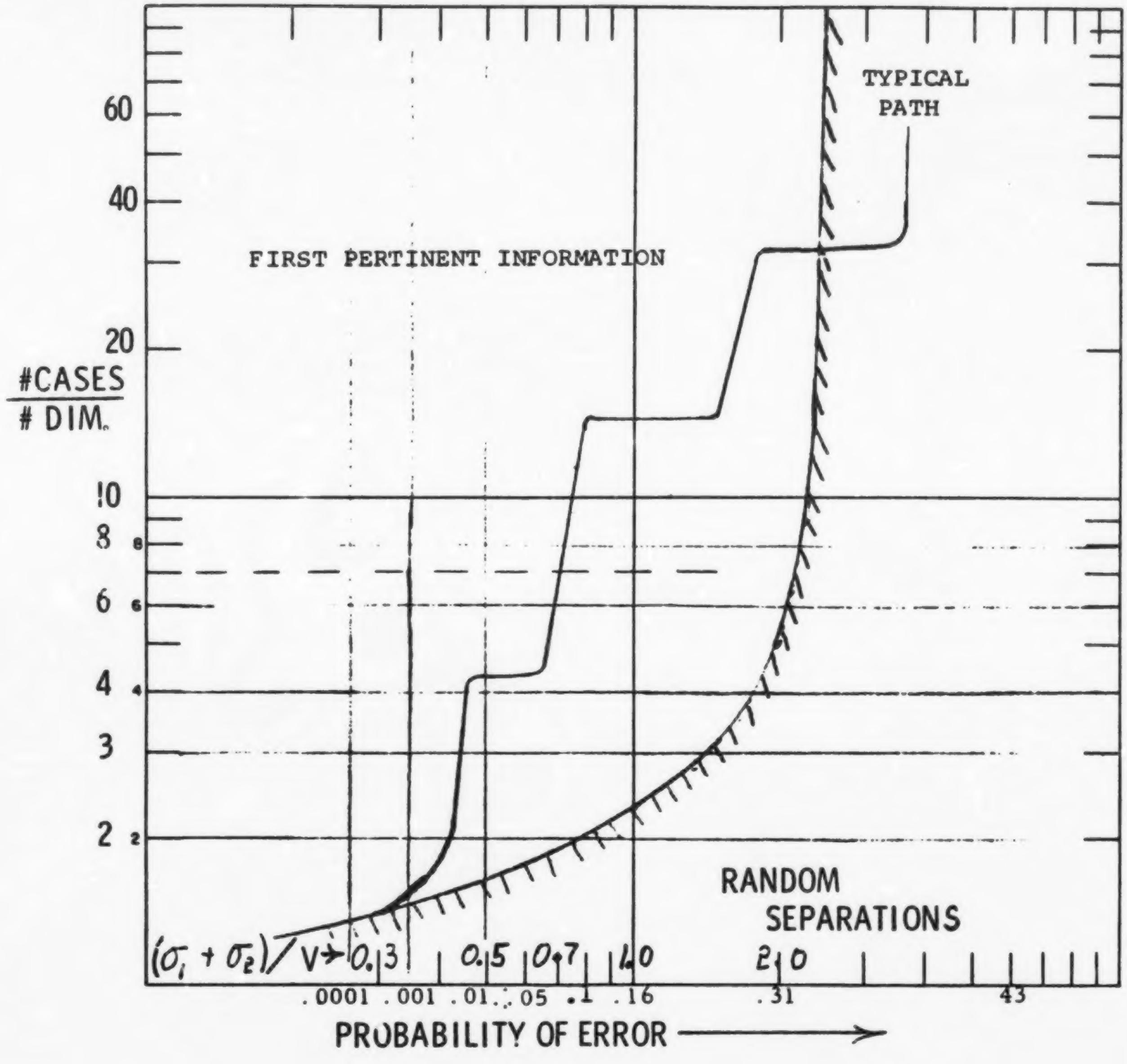
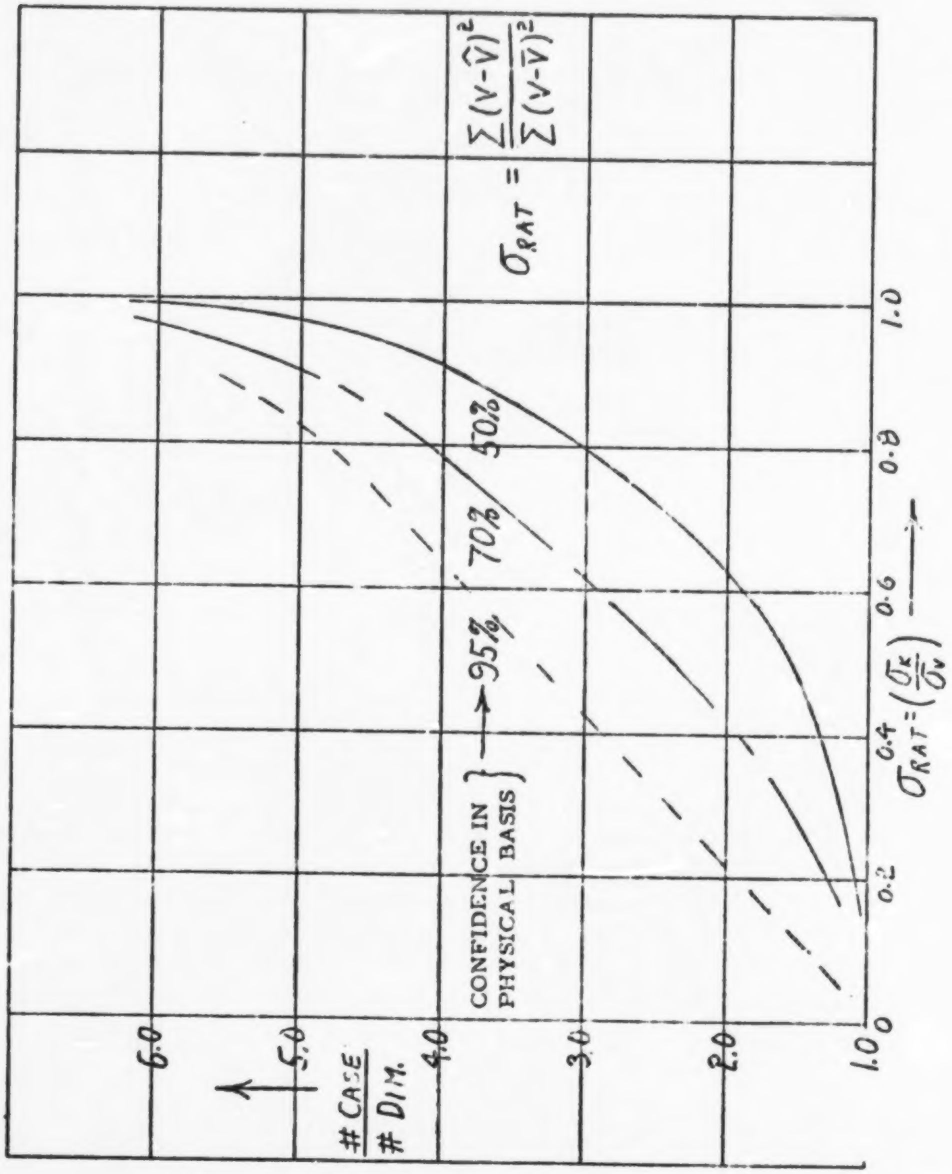


FIGURE - 14A
REGRESSION PERFORMANCE MAP



ADAPT REFERENCES

1A Watanabe, S., "Karhunen-Loeve Expansion and Factor Analysis Theoretical Remarks and Predictions", Transaction of the 4th Prague Conference on Information Theory, Statical Decision Functions, and Random Processes, 1965, pages 635 thru 660.

2A Andrews, Harry C., "Introduction to Mathematical Techniques in Pattern Recognition," John Wiley & Sons, Inc. 1972

3A Anderson, T.W. and Bahadur, R.R., "Classification into Two Multi-Variate Normal Distributions with Different Co-Variance Matrices" Annals of Mathematical Statistics, Vol. 33, p. 420, 1962.

4A Foley, Don H., "Probability of Error in the Design Set As A Function of A Sample Size Dimensionality", Thesis, Syracuse University, 1971.

5A Hunter, H.E., N. Kemp, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Cyclone Central Pressure and Its Derivatives Using NIMBUS HRIR Data", AVSD-0362-70-RR, August 1970.

6A Hunter, H.E., N. Kemp, "Application of Avco Data Analysis and Prediction (ADAPT) to Prediction of Cyclone Central Pressure and Its Derivatives Using NIMBUS HRIR Data", AVSD-0142-71-RR, March 1971.

7A Hunter, H. E., N. Kemp, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Cyclone Present Motion and 12 Hour Motion, and Re-centering Effects, Using NIMBUS HRIR Data", AVSD-0334-71-RR, July 1971.

8A Shenk, William E., Herbert E. Hunter, Frederick V. Menkello, Robert Holub, Vincent V. Salomonson, "The Estimation of Extra-tropical Cyclone Parameters from Satellite Radiation Measurements," Journal of Applied Meteorology April 1973.

9A Kemp, N.H., H. E. Hunter, R. A. Amato, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Multi-Spectral Extra-Tropical Cyclone Accuracy Investigation", AVSD-0128-72-CR, March 1972.

10 A Hunter, H. E., N. H. Kemp, "ADAPT Hurricane Data Selection and Performance Study", AVSD-0400-72-RR, Nov. 1972.

11 A Hunter, H. E., N. H. Kemp, "ADAPT Hurricane Forecast Improvement Demonstration", AVSD-0020-73-CR, January 1973.

12 A Hunter, H. E., "Use of Satellite Data and the ADAPT programs to Improve Hurricane Forecasts" AVSD-0138-73-RR, April 1973.

13 A Avco Corp., ADTECH (Advanced Decoy Technology) Program Final Report, Vol, 3, Appendices, Avco TR No. RAD TR-65-4, Contract AF04(694)-593, DDC #AD363081, April 30, 1965, (Secret) Pages 157-

14 A Avco Corporation , ADTECH II Final Report, May 1966, BSD-TR-66-192. Sponsored by Advanced Research Projects Agency (ARPA), DOD ARPA Order #441 Amendment #4, DDC #AD374278, May 1966. (Secret) Pages 64-89.

15 A Avco Corp., ADETCH III Final Report, Feb. 1968, AVMSD-0835-67-RR, Contract AF04(694)-956. (Secret)

16 A Avco Corp., ADTECH IV Final Report, June 1969, AVMSD-0465-68 RR, Sponsored by ARPA, DOD ARPA Order #441 Amendment #12 (Secret).

17 A Choiniere, Dill, Hines; Chaff Masking Effectiveness Paper 61 in AMRAC Proceedings, Volume XVIII, Part I, (AD-390700), Published by University of Michigan, April 1968. (Secret)

18 A Avco Corp., Test and Evaluation Study Report - Vol. II, Data Bank Study, Prepared for Institute for Defense Analysis, Contract FO 4701-68-C-0012, AVMSD-0300-68-RR, 22 April 1968. (Secret)

19 A Hunter, H. E., "Discussion of Patterns Recognition Techniques Applied to Diagnosis", presented at the Society of Automotive Engineers, Mid-year meeting, May 18-20, 1970.

20 A Hunter, H. E., R. Amato, J. Conway, N. Kemp, "Demonstration of Applicability of Avco Data Analysis Technique to Sonar Signature," AVSD-0605-70-RR, December 1970.

21 A Hunter, H. E., J. Conway, "Demonstration of Feasibility of Using the Avco Data Analysis and Prediction Techniques (ADAPT) to Develop Algorithms for Automating the Identification of Solar Burst", AVSD-0255-71-RR, 21 May 1971.

22 A Avco Corp., Tethered Radar Reflectors (TRR) Report, Sept. 1971. SAMSO TR-71-181, Vol. II, Contract F04701-68-C-0289. (Secret)

23 A Hunter, H. E., N. Kemp, "Demonstration of Feasibility of Avco Data Analysis and Prediction Techniques (ADAPT) for Sonar Detection" AVSD-0411-71-RR, September 1971.

24 A Hunter, H. E., L. Meixsell and J. Conway, "Feasibility Demonstration for Optically Implementing an Avco Data Analysis and Prediction Techniques (ADAPT) Algorithm for Recognizing Spirals", AVSD-0026-72-RR, January 1972.

25 A Hunter, H. E., J. Conway, "Use of the Avco Data Analysis and Prediction Techniques (ADAPT) to Develop Analytical Techniques for a Comprehensive Attach on Auto Theft and Burglary in Lawrence, Mass.", AVSD-0042-72-RR, 31 January 1972.

26 A Hunter, H. E., L. M. Meixsell, "Preliminary ADAPT Analysis Feasibility of Discriminating Crash Sensor Signature," AVSD-0398-71-RR, 30 Aug. 1971.

27 A Hunter, H. E., "Application of ADAPT to Analysis of Bi-Static RCS Crash Signature," AVSD-0180-72-RR, May 1972.

28 A Jones, T. O., D. M. Grimes, R. A. Dork, "A Critical Review of Radar as a Predictive Crash Sensors," presented at the Second International Conference on Passive Restraints, Detroit, Michigan, May 22-25, 1972, SAE Report 720424, Pages 22-24, 38.

29 A Hunter, H. E., L. M. Meixsell, R. A. Amato, "Final Letter Report ADAPT Solar Burst Compacting Study," AVSD-0209-72-CR, May 1972.

30 A Kemp, N. H., H. E. Hunter, R. A. Amato, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to a Gauss-in-Gauss Detection Study, " AVSD-0260-72-RR, July 1972.

31 A Hunter, H. E., R. A. Amato, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Sunspot Activity," AVSD-0287-72-CR, August 1972.

32 A Hunter, H. E., "Demonstration of the Use of ADAPT to Derive Predictive Maintenance Algorithms for the KSC Central Heat Plant," Nov. 1972, Contract No. NAS 10-7926, AVSD-0084-73-RR.

33 A Kemp, N. H., H. E. Hunter, "Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Analysis of TUMS Sonar Data," AVSD-0433-72-CR, December 1972.

34 A Hunter, H. E., "Application of ADAPT to Determination of Effect of Diesel Capsule Valve Design Criteria on Fuel Flow Performance," AVSD-0102-73-RR, March 1973.

35 A Kemp, N.H., H. E. Hunter, R.A. Amato, "Avco Data Analysis and Prediction Techniques (ADAPT) Tri-Class Passive Sonar Classification Study, January 1973, (Confidential)

36 A Hunter, H.E. "Application of ADAPT to Selecting Optimal Features for Study and Modeling of Rain Cell Radar Signatures" AVSD-0122-73-RR, April 1973.

37 A Hunter, H.E., "Final Report-ADAPT Cyclone Forecast Correction Study", ADAPT 73-1, September 1973.

38 A Hunter, H.E., "Task-1 Final Report - Application of ADAPT to Quick Look Classification of Composite Radar Signatures", ADAPT 73-3, November 1973.

39 A Hunter, H.E., "Final Report-Application of ADAPT to Integrated Trend Analysis for Checkout of Space Vehicles", ADAPT 74-2, April 1974

40 A Hunter, H.E., "Final Report-Application of ADAPT to Quick Look Classification of Composite Radar Signatures", ADAPT 74-4, ESD TR-74-345, November 1974.

41 A Hunter, H.E., "Summary Letter Report - ADAPT Studies to Define Diagnostic Potential of Preliminary Brake Analysis Data", ADAPT 75-1, May 1975.

APPENDIX B

ANALYSIS OF OPTIMAL BASES

Two ADAPT optimal bases were prepared for the analysis of the Turkey Point salt concentration data. These optimal bases were prepared using the procedures described in Appendix 1. Originally only one optimal base was to be prepared, however, analysis of this base showed that there were significant keypunching errors in its preparation. It was necessary to prepare a second base using the corrected data. This appendix presents the most significant characteristics of each of these bases. These characteristics are presented as scatter plots, plots of the ADAPT optimal functions and ADAPT information energy plots. For the interpretation and meaning of each of these plots, the reader is referred to the descriptions provided in Appendix A.

Fig 1-B presents the scatter plot of every third case from the first two optimal coordinates for the original base. This scatter plot led to the discovery of the keypunching errors in the data. This scatter plot is the projection of each of the data histories on the first two ADAPT optimal directions. Both the data histories and optimal directions are made up of Variables 3 through 79 listed under the 82pt column of Table 1 in the main body of the report. This projection is obtained by taking the dot product of each of these data histories with the ADAPT optimal function corresponding to the coordinate onto which the data history is to be projected. Thus, the abscissa of Fig 1-B is obtained by taking the dot product of the data history with the first ADAPT optimal function presented in Figure 2B. The ordinate of Figure 1-B is obtained by taking the dot product of the data history with the second ADAPT optimal function which is presented in Figure 3B. Thus, the ADAPT optimal functions may be considered as relative importance vectors for defining the location of a point on the scatter plot coordinate corresponding to that optimal function. Examination of Figure 2B shows that the first optimal function is primarily a time measure such that projections on the first optimal function should have negative values for tests performed early in the program and positive values for tests performed later in the program. Examination of the second optimal function presented in Figure 3B shows that the second optimal function is primarily a measure of whether the cooling device was operating or not.

FIGURE 1B- PROJECTION OF EVERY THIRD CASE ON FIRST TWO OPTIMAL COORDINATES OF FIRST ANALYSIS BASE

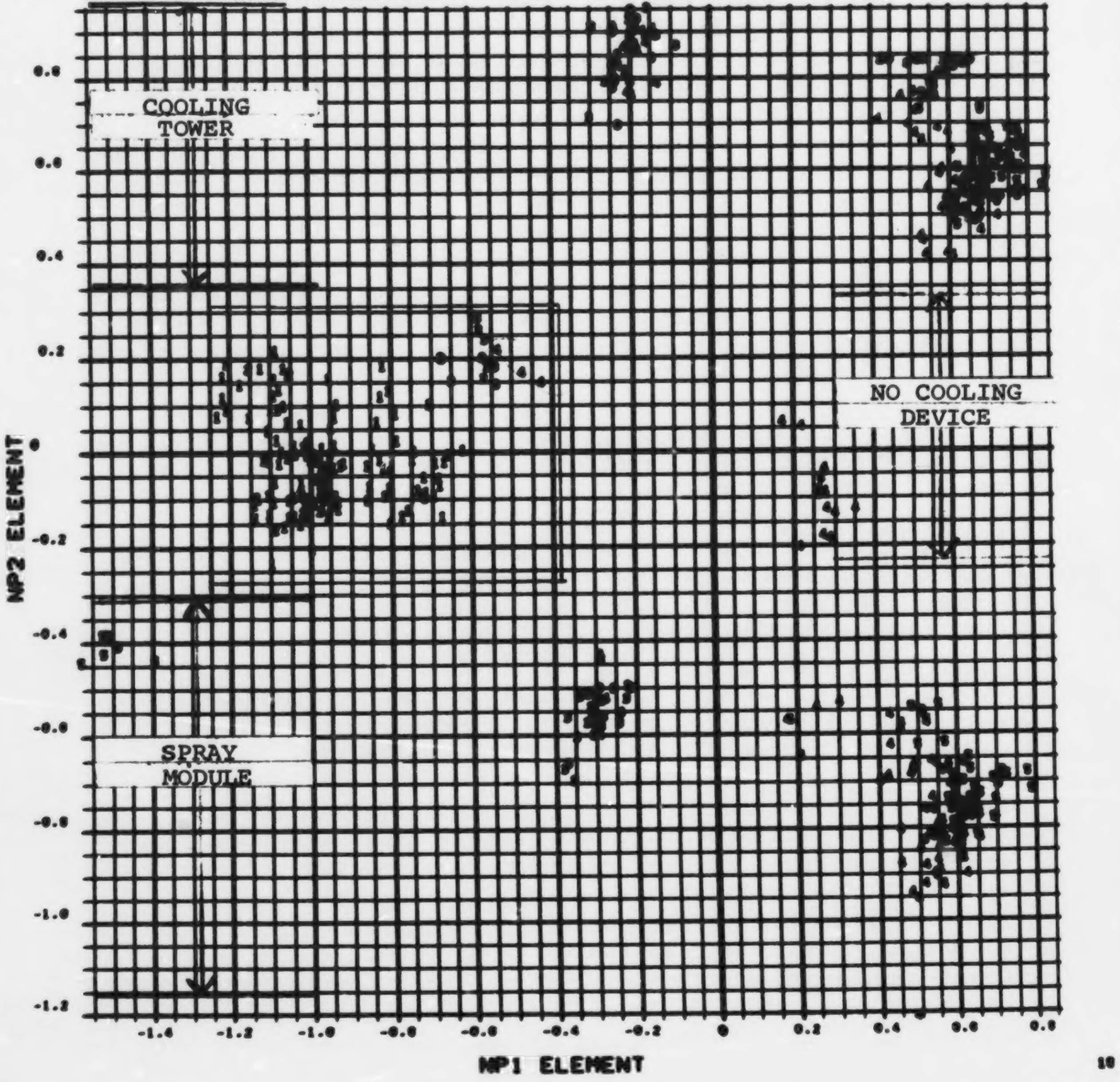


FIGURE 2B
FIRST ADAPT OTPIMAL FUNCTION FOR FIRST ANALYSIS BASE

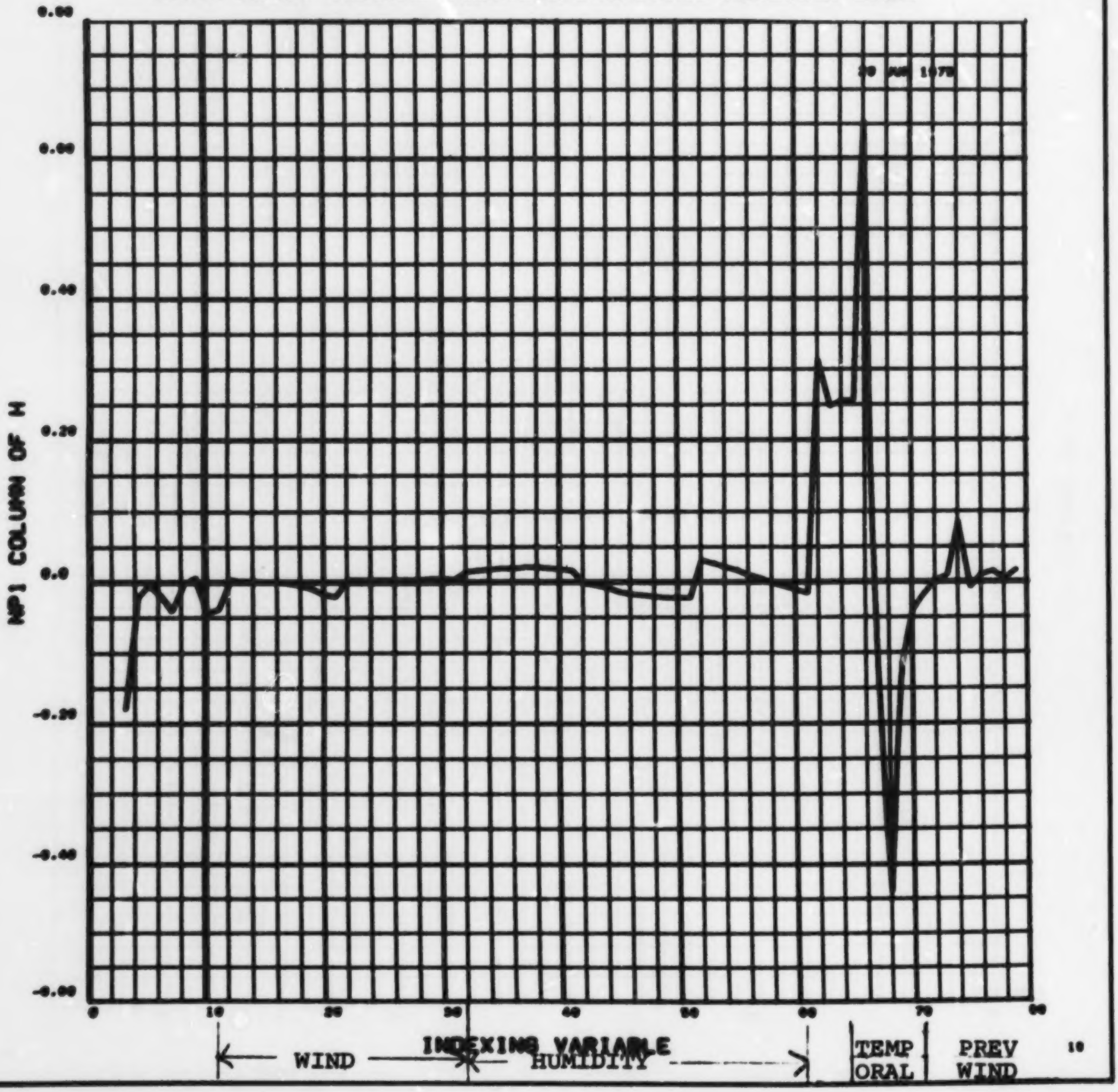
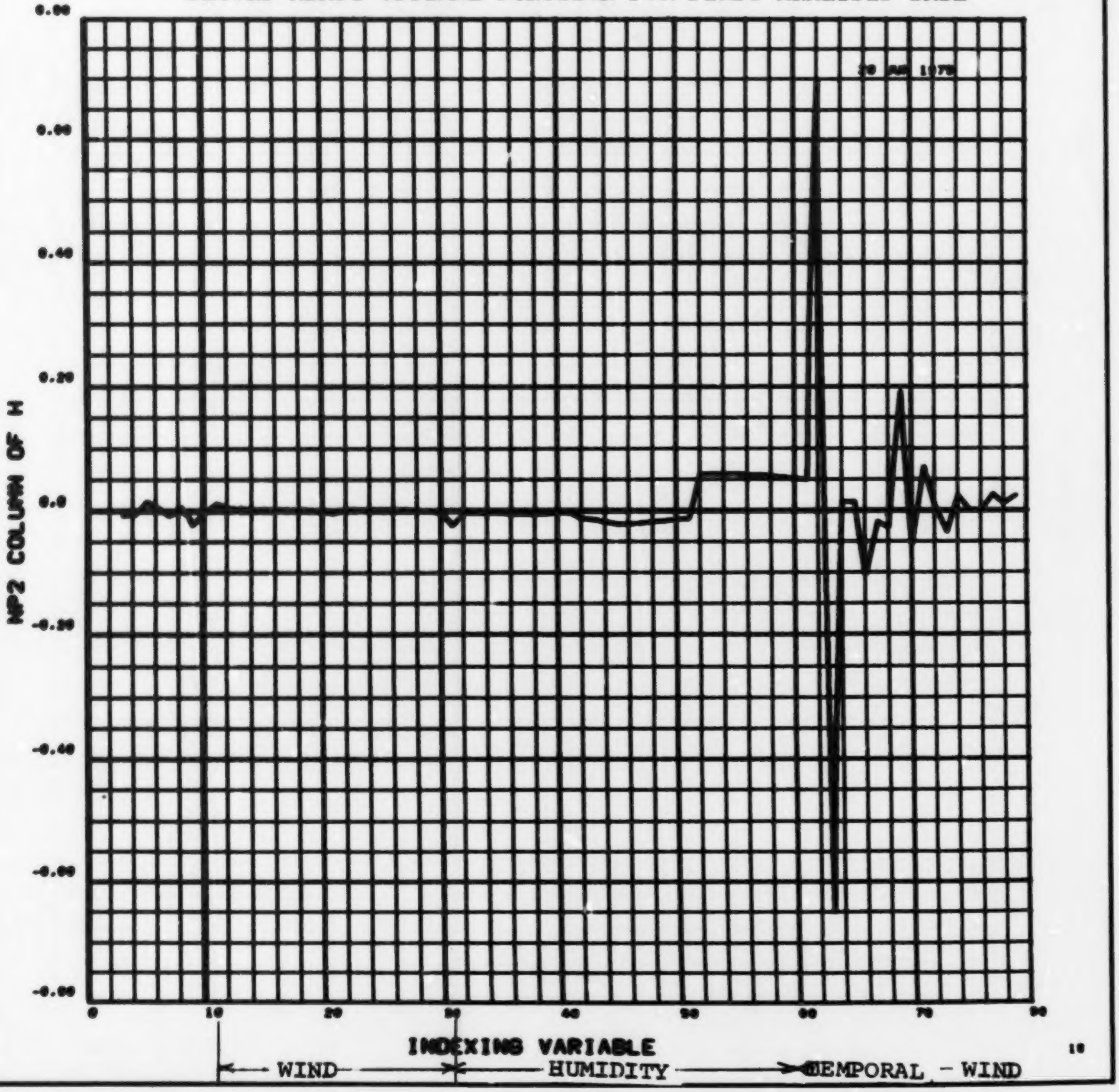


FIGURE 3B
SECOND ADAPT OPTIMAL FUNCTION FOR FIRST ANALYSIS BASE



High values of the projectional on the second optimal function correspond to cases where the cooling tower was operating, values near zero correspond to the time period where no cooling device was operating and large negative values correspond to the time values where the spray module was operating.

The numbers 1 through 5 used to designate the data histories projected on the scatter plot shown in Figure 1-B are a chronological ordering of the data histories. The Number 1 designates those data histories which were obtained during the time period when neither cooling device was operating. The Numeral 2 designates those cases obtained during the remainder of 1973. The 3's and 4's are from early 1974 and the 5's are from the later period of 1974 when data were obtained with a cooling device operating. Examination of Fig 1-A shows the anomolous result that a group of 5's are located on the left hand side with an NP1 projection of minus 1.4 to minus 1.5. This is inconsistent with the definition of the first ADAPT optimal function presented in Fig 2-B since the fives should occur on the right hand side of this figure. There are also several anomoulous groupings of numberal 4's on this scatter plot. Investigation of each of these groupings show that they were a result of keypunching errors in the data preparation. Since these keypunching errors had a significant effect on the variation of the data set used to derive the optimal base, it was necessary to recreate this base using the corrected data.

The average of all of the data histories used to develop the second analysis base using the corrected data is shown in Fig 4-B Prior to processing these data histories to develop this ADAPT optimal base, this average is subtracted from all the data histories so that all of the succeeding analysis is performed on zero mean data. Fig 5-B presents a plot of the effect of dimensionality on the information available for analysis using the second analysis base. Fig 5-B is actually the plot of two curves on a single grid. The lower curve presents the amount of information available in each of the terms of the optimal base. The upper curve is the cumulative sum of the lower curve. Thus, this curve indicates that the scatter plot containing the first two dimensions in this optimal base contains approximately a third of the information in the entire set of data being analyzed. Similarly, if one performs an analysis using 16 dimensions the sixteenth dimension

FIGURE 4B - AVERAGE OF ALL DATA HISTORIES USED TO DEVELOP SECOND ANALYSIS BASE

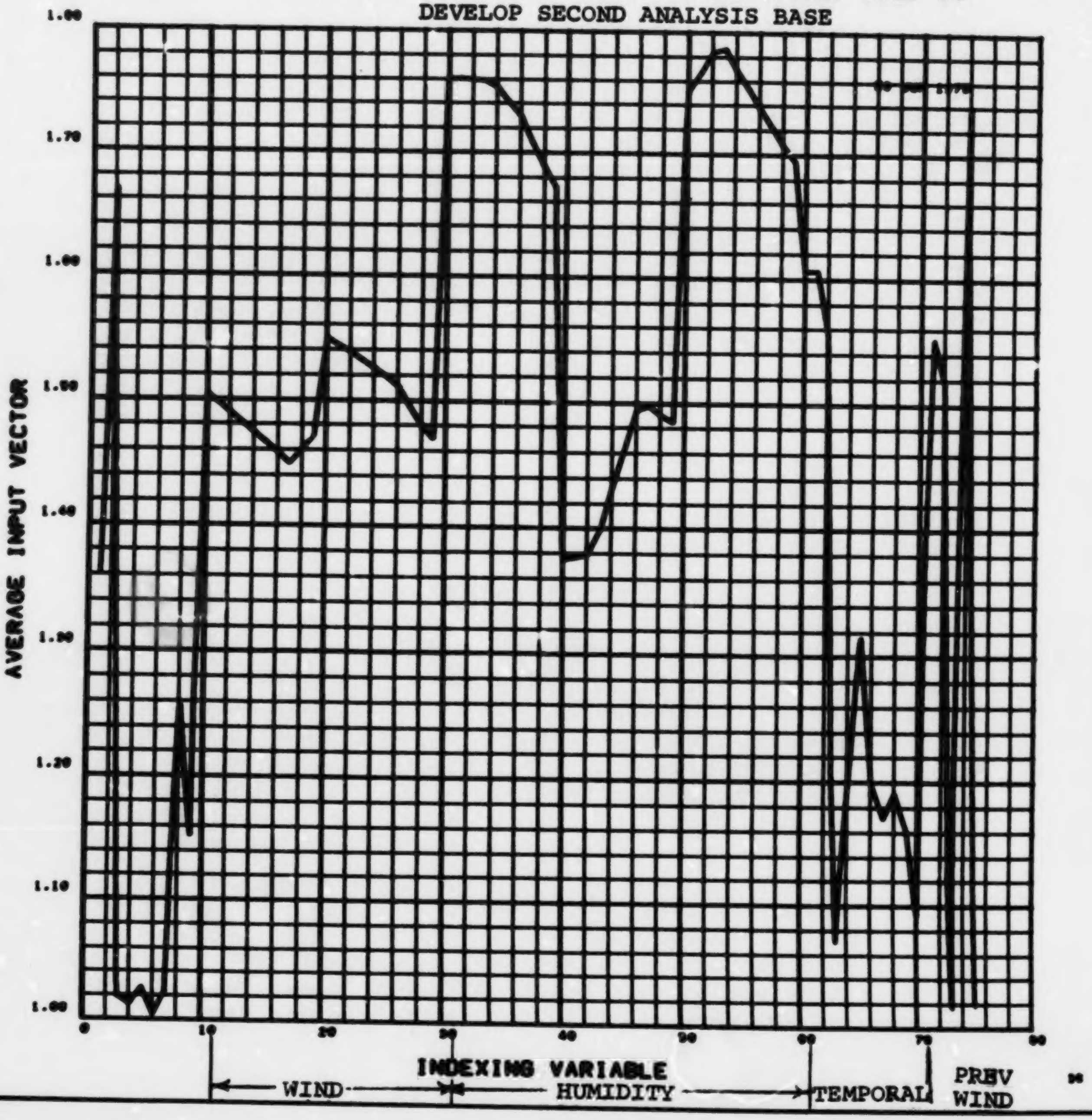
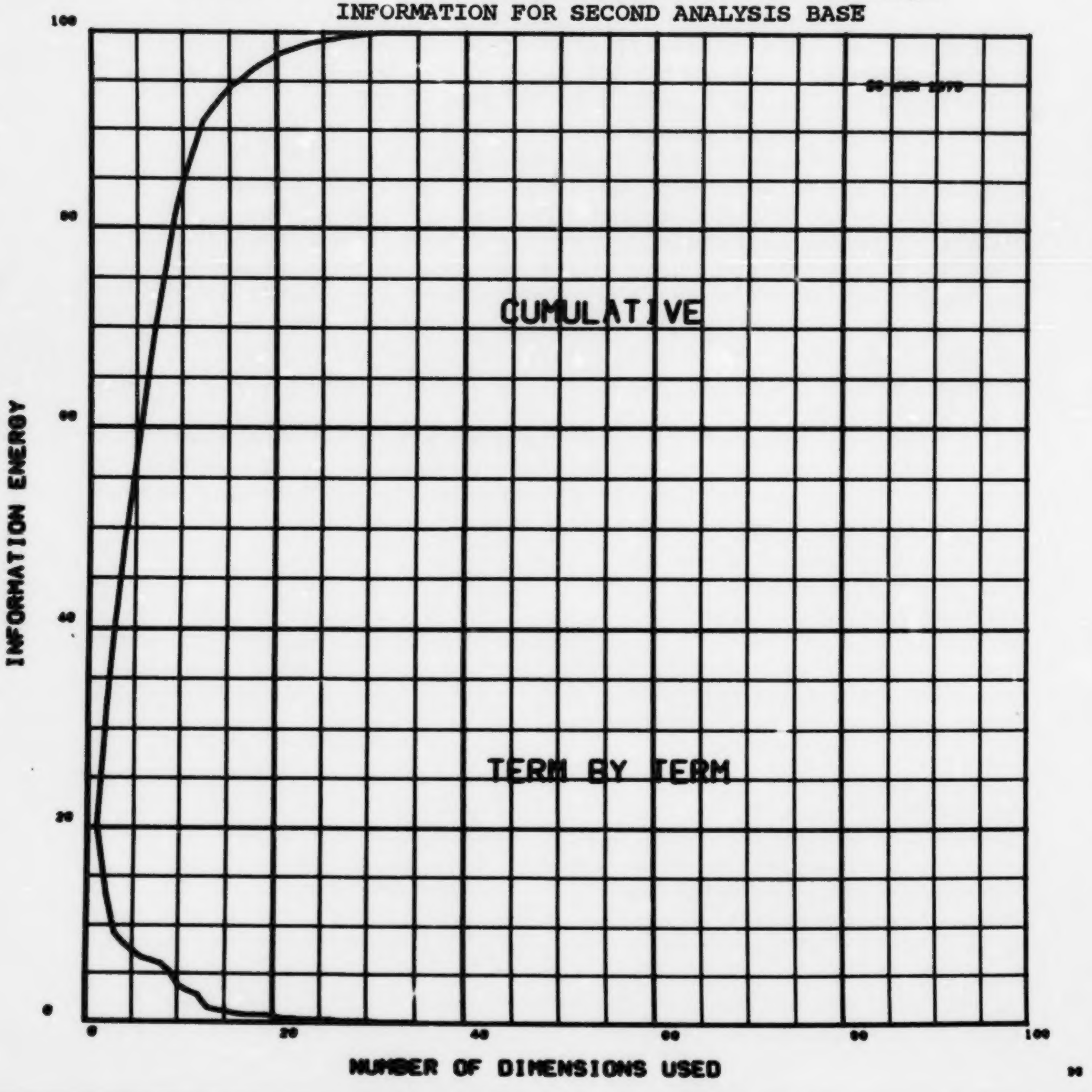


FIGURE 5B- EFFECT OF DIMENSIONALITY ON THE AVAILABLE INFORMATION FOR SECOND ANALYSIS BASE



only contains approximately 1% of the information contained in the original data set. The first sixteen dimensions taken together contain approximately 95% of the information available in the entire data set. Analysis of the optimal functions associated with this base suggested that up to approximately the first 26 optimal dimensions the information contained was sufficiently general to be useful to the type of analysis being performed in this study. However, the number of cases available for analysis restricted the dimensionality to between 4 and 20 dimensions depending upon a particular algorithm being developed. This implies that almost 99% of the information available in the data set could be useful to the present analysis. Fig 5-B shows that the limitations on the number of cases available for analysis have restricted the amount of information which could be used in this study to approximately 40% for the four dimensional algorithms to as much as 98% for the 20 dimensional algorithms. The majority of the algorithms were developed at 16 dimensions which corresponded to approximately 95% of the available information.

Fig 6-B presents the scatter plot projection of every third case on the first two optimal dimensions of the second analysis base. The effect of correcting the keypunching errors can be seen by comparing Figures 1B through 3B with Figures 6-B through 8B. Fig 7-B and 8B present the first and second optimal functions using the corrected data. The first optimal function remains strongly dependent on the temporal variables, however, humidity and wind are more important to the first optimal function then in the base with the keypunch errors. The second optimal function is considerably different and is dominated by information concerning the wind and humidity.

The numerals used to designate the bases on Figure 6-B have the same meaning as for Figure 1B. Examination of Figure 6-B shows that in general the numeral indicating the test increases from left to right on this figure. This is in general an agreement with the temporal nature of the first optimal function. Note, that there are two major groupings of data on this scatter plot. This indicates there are probably significant differences between the characteristics of the tests performed in the first half of the program from those performed in the second half of the program. These groupings indicate potential areas of incomplete data. For this data set, the majority of the scatter plots are similar to the scatter plot projection data on the third and fourth optimal directions shown in Figure 9B.

FIGURE 6B- PROJECTION OF EVERY THIRD CASE ON FIRST TWO OPTIMAL COORDINATES OF SECOND ANALYSIS BASE

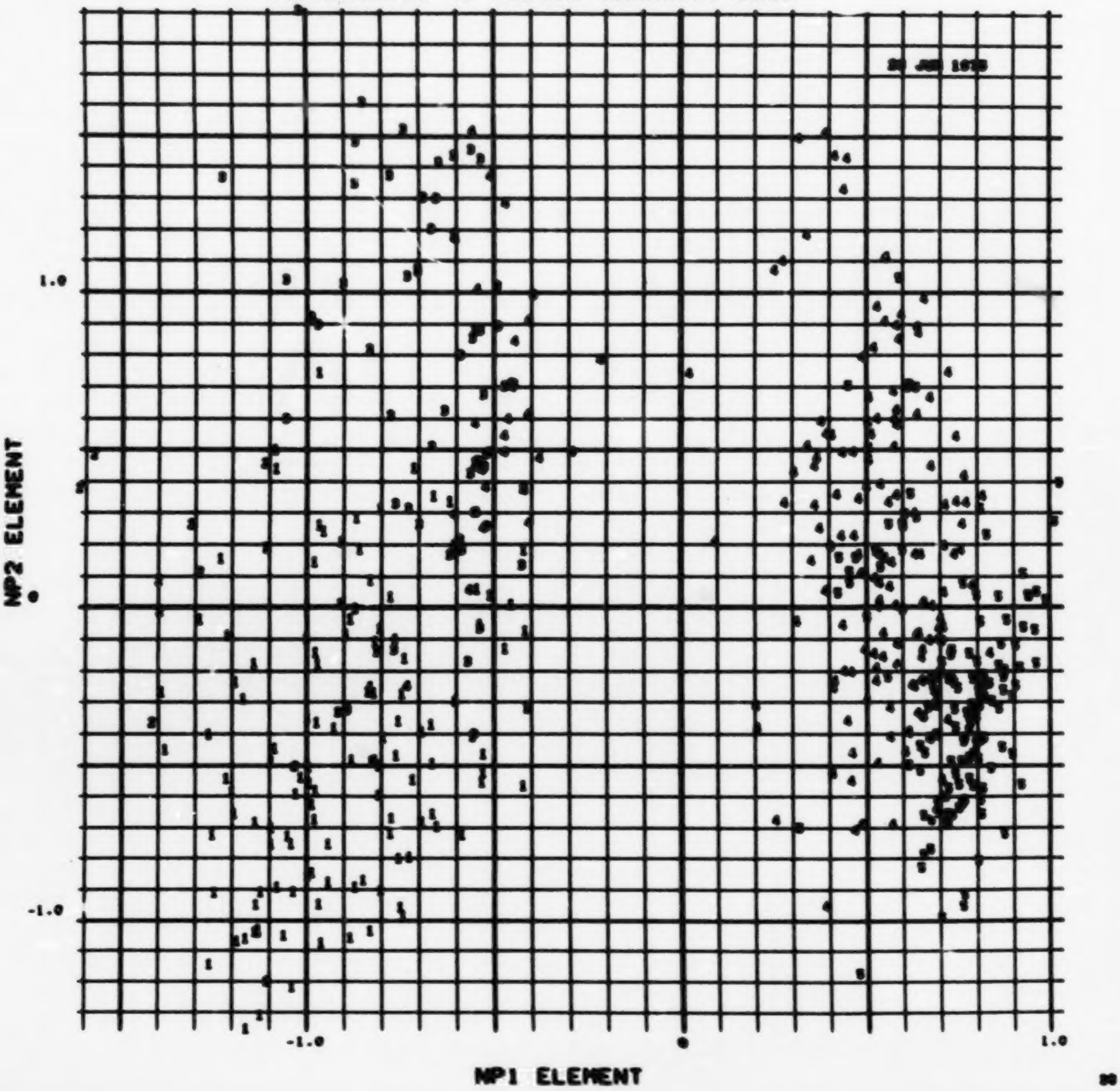


FIGURE 7B. - FIRST OPTIMAL FUNCTION SECOND ANALYSIS BASE

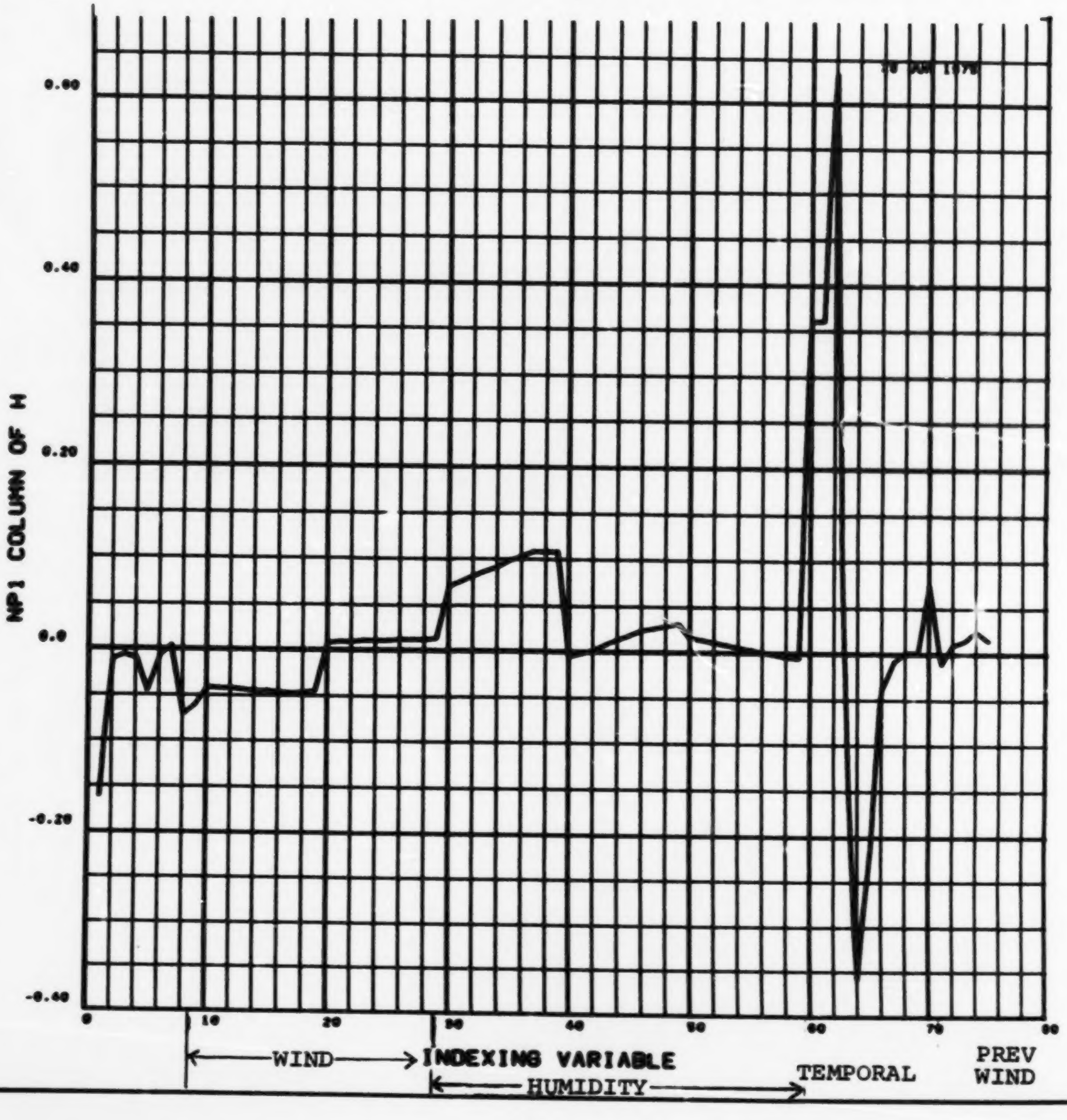


FIGURE 8B - SECOND OPTIMAL FUNCTION SECOND ANALYSIS BASE

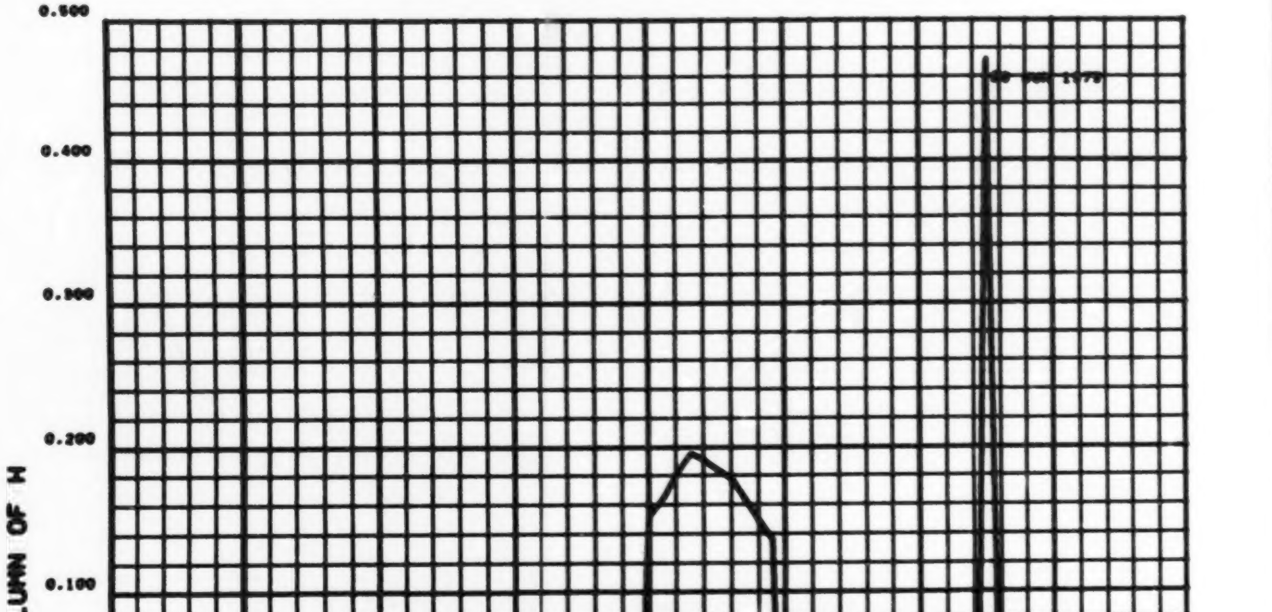
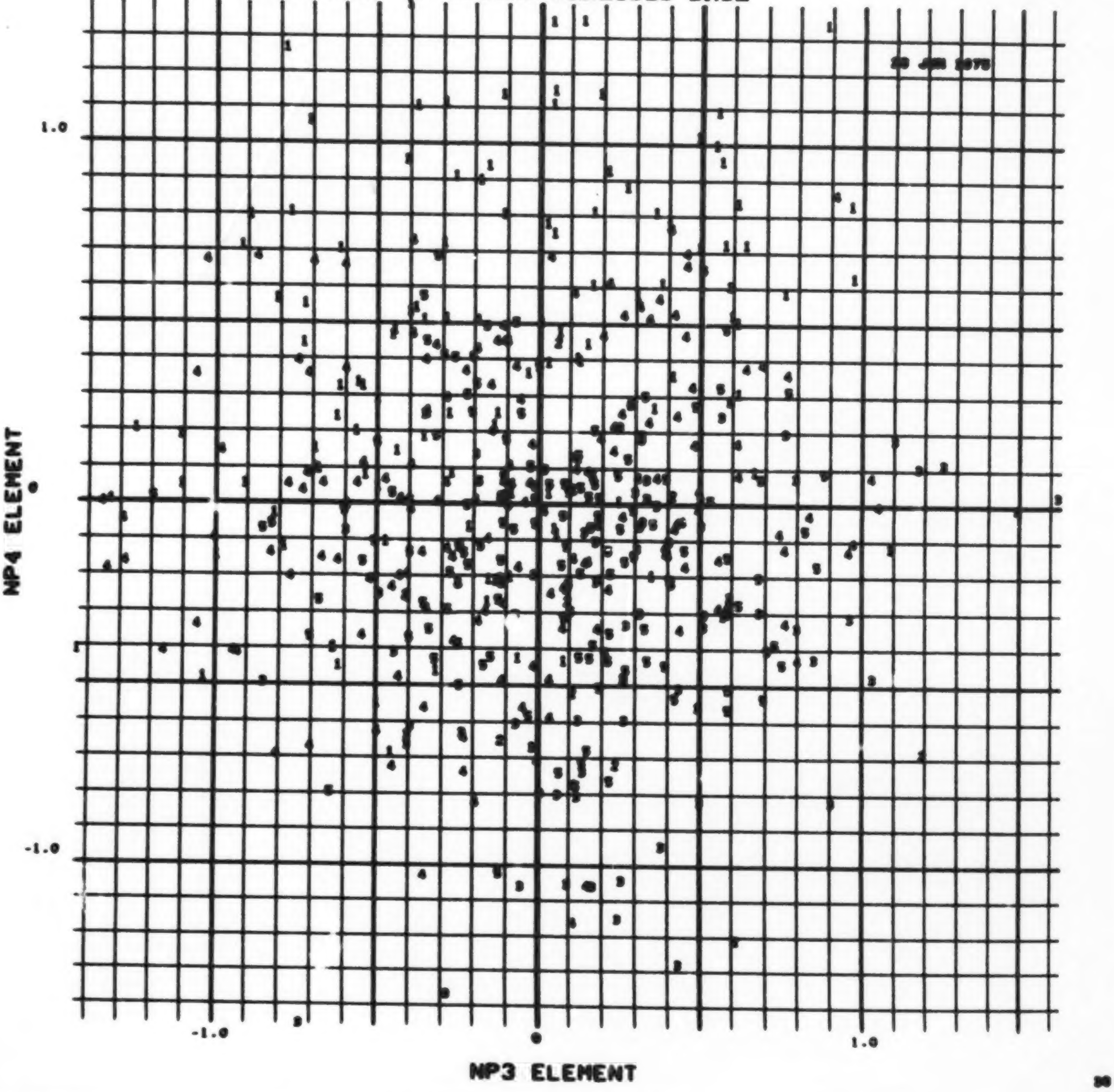


FIGURE 9B- PROJECTION OF EVERY THIRD CASE ON THIRD AND FOURTH OPTIMAL DIRECTIONS OF SECOND ANALYSIS BASE



This scatter plot shows an even distribution of the data indicating that the data set should be complete for the analysis over the variation in variables considered for this study. The next significant dichotomy occurred in the scatter plot of the seventh and eighth optimal functions. This scatter plot is shown in Figure 10B. The grouping in the lower right hand corner of this scatter plot may be attributed to cases for which the cross wind values are low. This follows from the seventh optimal function which is shown in Figure 11B. This function is dominated by the cross wind variables, 20 through 29.

FIGURE 10B - PROJECTION OF EVERY THIRD CASE ON SEVENTH AND EIGHTH OPTIMAL DIRECTION OF SECOND ANALYSIS BASE

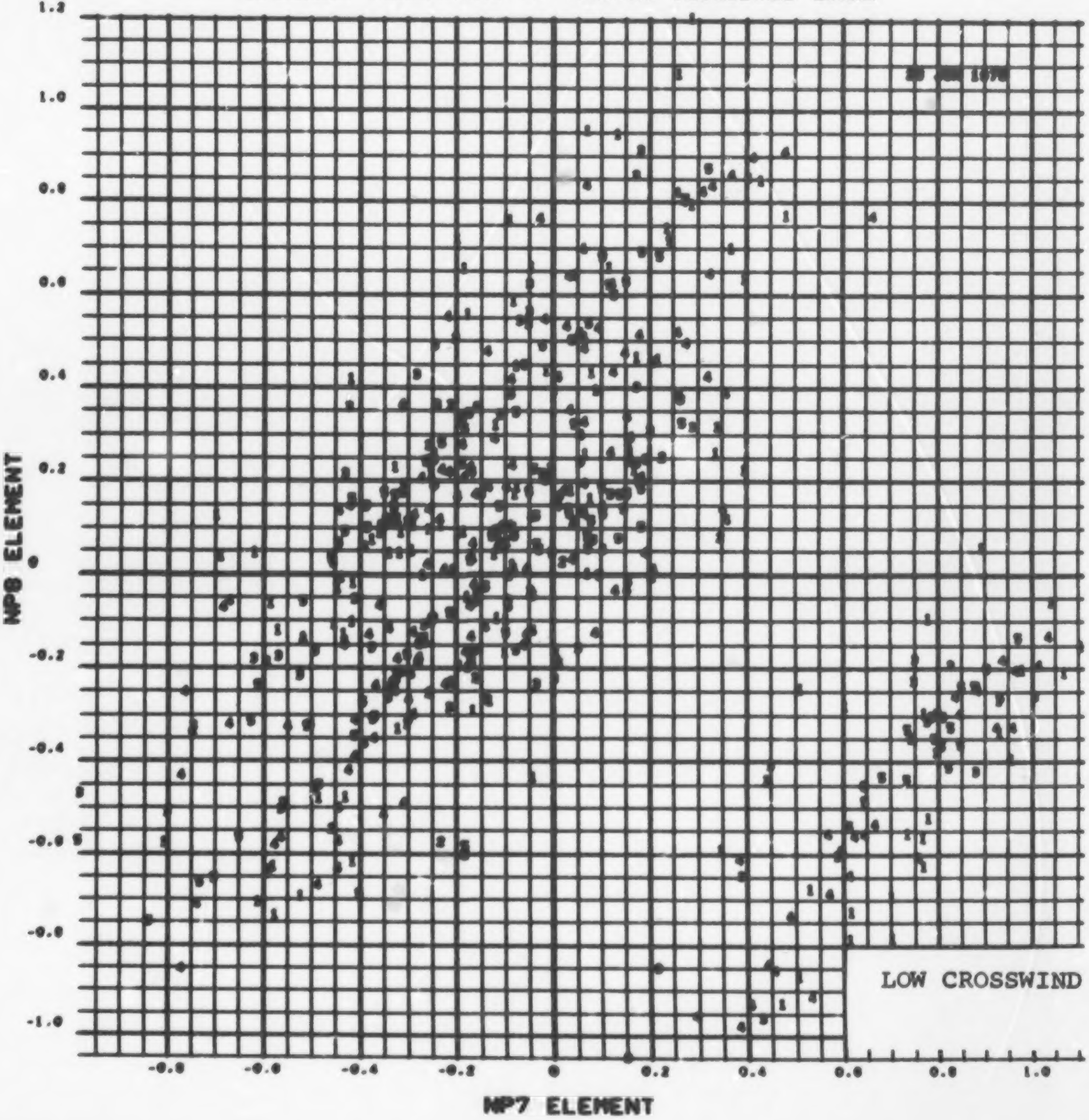
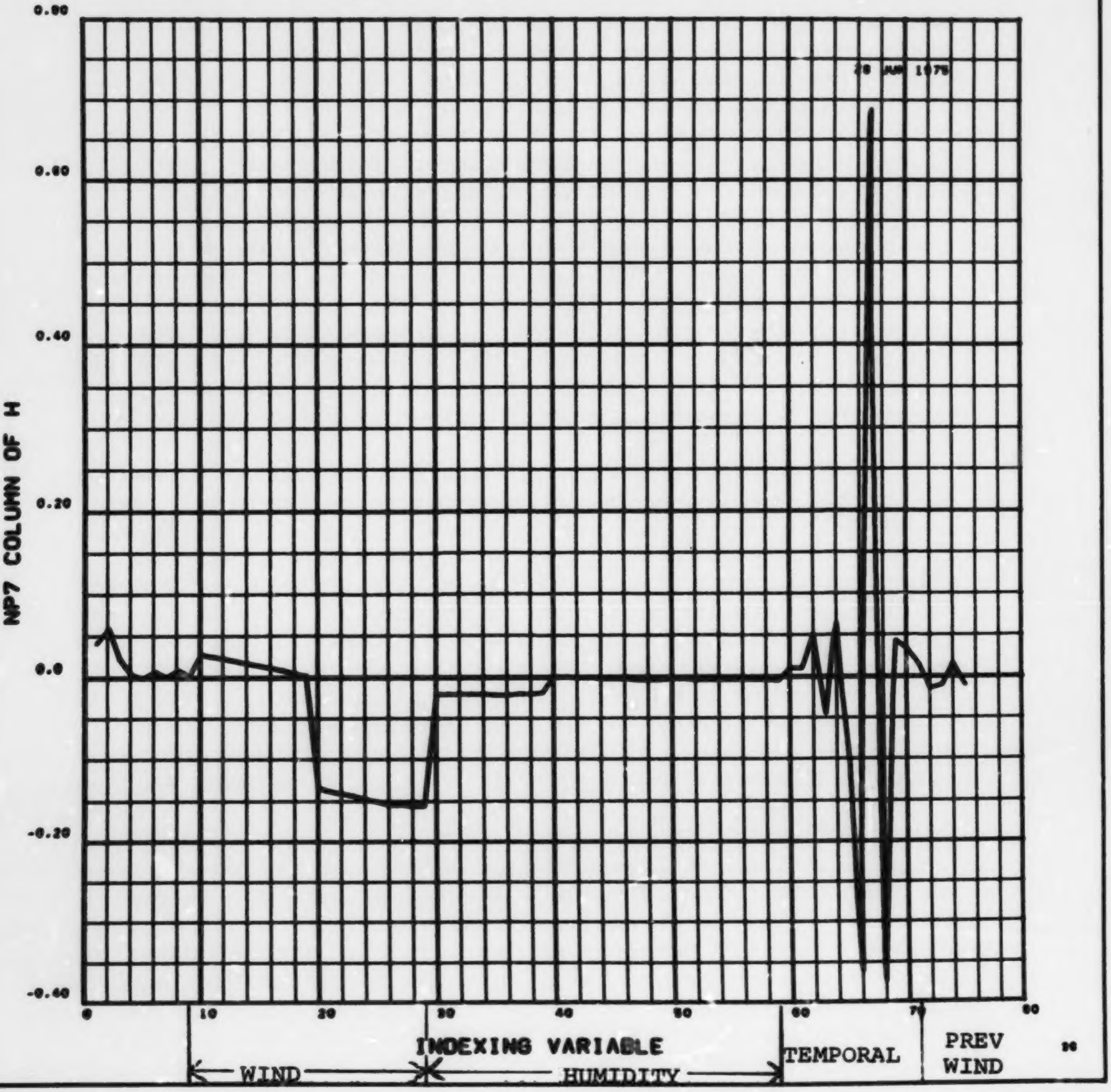


FIGURE 11B - SEVENTH OPTIMAL FUNCTION SECOND ANALYSIS BASE



APPENDIX C

SELECTION AND ANALYSIS OF ALGORITHMS

Two types of algorithms were developed for the present study. The first was an algorithm to estimate the precision run error as a function of the environment and the second were algorithms to estimate the background concentration as a function ofthe environment and station location.

The algorithm for estimating the precision run error was developed using available variables which defined the environment at Stations 1 and 2. Unfortunately, only 65 cases were available to develop this algorithm. Analysis of the information energy indicated that one should use more than 20 dimensions for this algorithm. However, the 65 cases restricts one to less than 11 dimensions. Algorithms were developed using both 20 and 11 dimensions. Confidence in applying these algorithms to independent test cases is low because of the dimensionality required by the poor ratio of the number of independent cases to the dimensionality of the analysis. The relative importance vectors for the 11 dimensional analysis is presented in the main report. The ADAPT performance map suggests that the relative importance vector for the 20 dimensional analysis is not meaningful because of the small number of cases available.

Several options were available for algorithms to estimate the background concentration at each of the stations. The first option was to use the data pooled over all stations to make an algorithm which would estimate the background concentration as a function of both location and environment. As discussed in the main body of the report, this option has the severe restriction imposed by the manner in which the wind vector was defined. Thus, it was decided that for this study one would develop algorithms to estimate the background concentration at each of the individual stations. These algorithms determine the effect of the environment on the background concentration of each of the measurement stations.

At all stations having approximately 100 or greater measurements available, algorithms were developed using 20, 16 and 12 dimensions. The performance of these algorithms was then evaluated using the ADAPT performance map. In all of these cases, the ADAPT performance map indicated that the 20 dimensional analysis should yield physically meaningful algorithms. This implies that the relative importance vector for the 20 dimensional analysis should have physically significant meaning. However, the validity of the algorithm when applied independent test cases may be limited to a small set of cases for the higher dimensional algorithms. For many applications, the ADAPT validity criteria (See appendix A) can be used to eliminate those cases for which the algorithm is not valid. However, for the present study the amount of data available was not sufficient to allow one to discard a significant portion of the learning cases to satisfy the validity criteria. Thus, a technique for selecting a dimensionality which would insure the applicability of the algorithms to almost all of the cases was required.

The technique used to select a dimensionality allowing sufficient generality of the resulting algorithms consisted of comparing the average value estimated for the station under the environmental conditions during which the cooling device was operating with the average value observed at the station during the time which the cooling device was not operating. A significant difference between these two values suggests that the dimensionality used was too high and restricted the applicability of the algorithm to a set of cases which did not include a significant portion the environmental conditions occurring during the actual operation of the cooling device. For the cases where more than 100 learning cases were available, 16 dimensions were ususally adequate. The exceptions to this were Stations 7 and 10 where only 12 dimensions could be used. The algorithms developed for Stations 1 and 2 were developed using 20 dimensions since these algorithms were not used to estimate background concentrations. They were only used for interpretation of the relative importance vectors. Since significantly less cases were available at Stations 8, 9 and 11, the dimensionalities of 4, 8 and 4 were used, respectively. The relative importance vectors for the algorithms using these reference dimensionalities are presented in the main body of the report for Stations 9 and 10 and in Figure 1-C through 8-C of this appendix for the remainder of the stations. These relative importance vectors were used to develop Tables 7 through 16 in the main body of the report.

FIGURE 1C

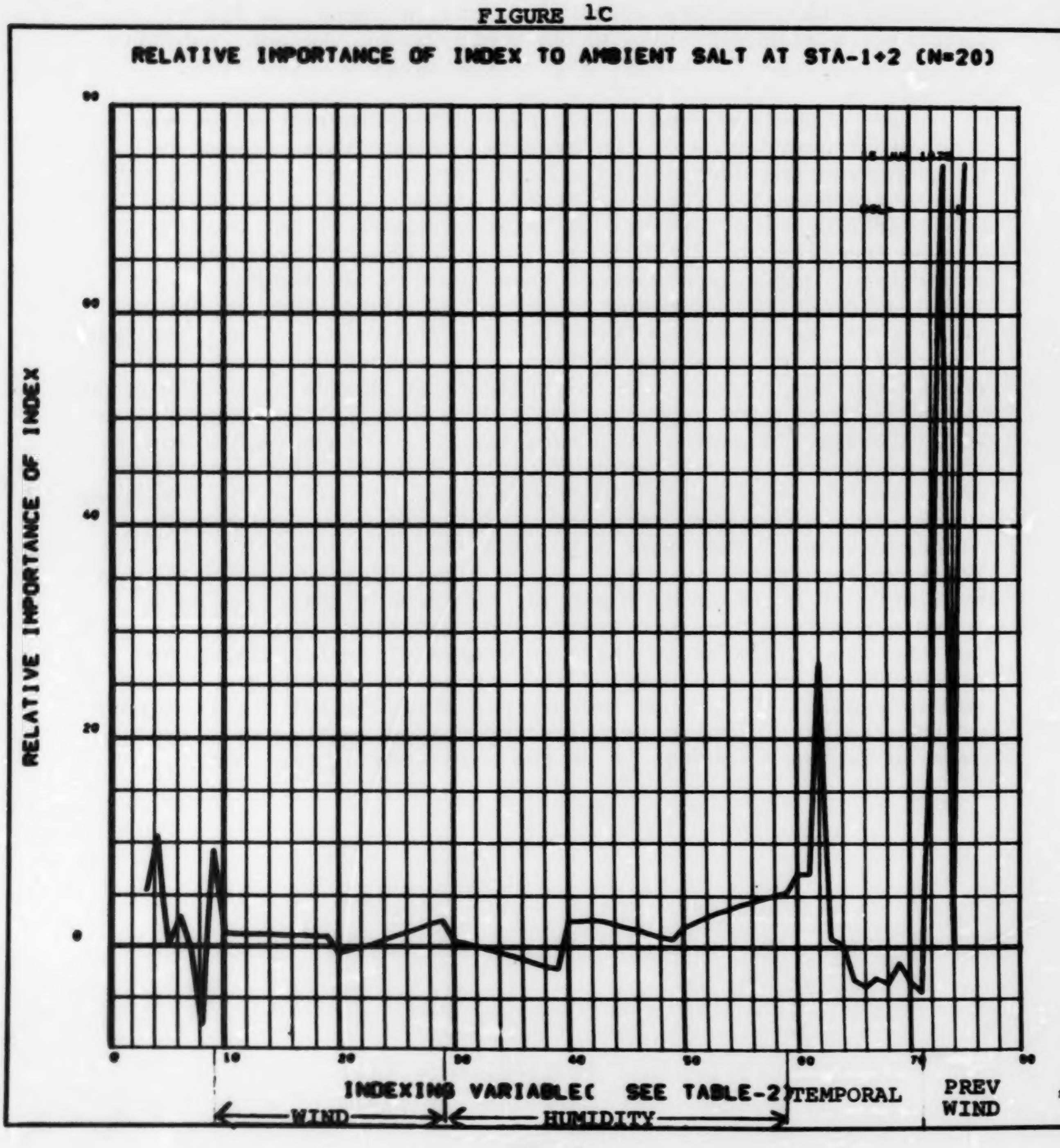

RELATIVE IMPORTANCE OF INDEX TO AMBIENT SALT AT STA-1+2 (N=20)

FIGURE 2C

RELATIVE IMPORTANCE OF INDEX TO AMBIENT SALT AT STA-3 (N=16)

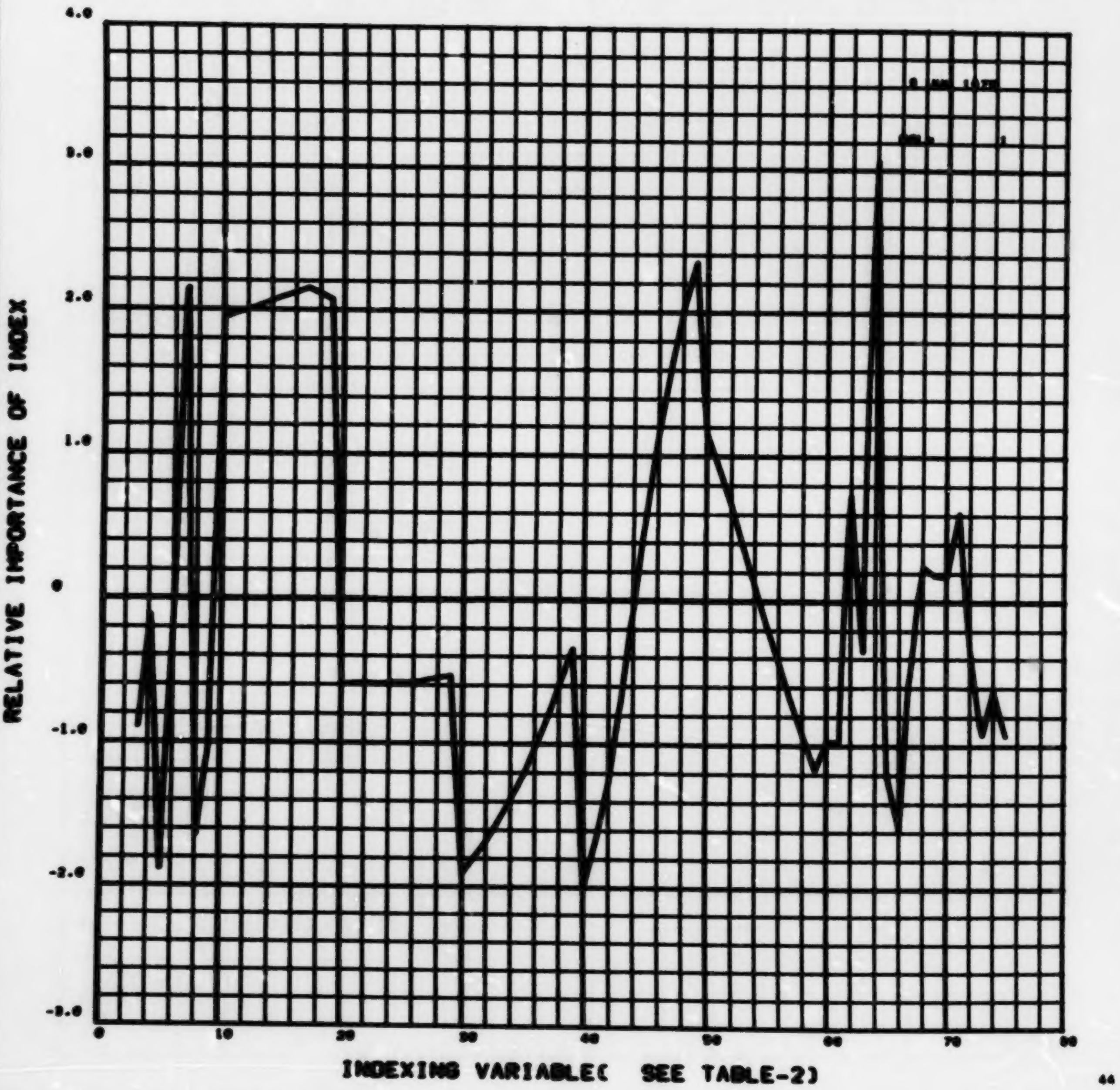


FIG- 3CRELATIVE IMPORTANCE VECTOR FOR AMBIENT CONC AT STATION-4 (N=16)

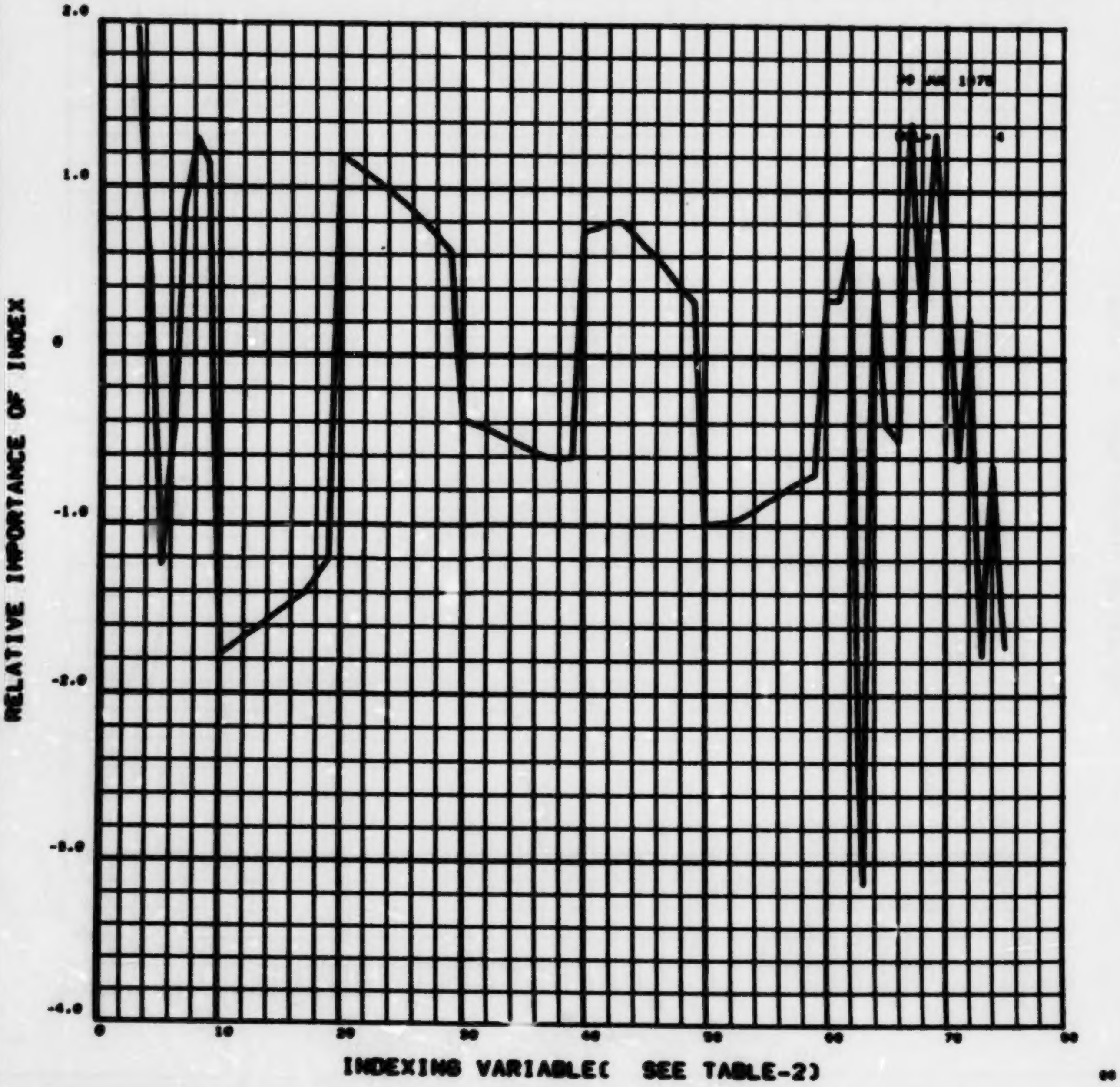


FIG- 4C RELATIVE IMPORTANCE VECTOR FOR AMBIENT CONC AT STATION-5 (N=16)

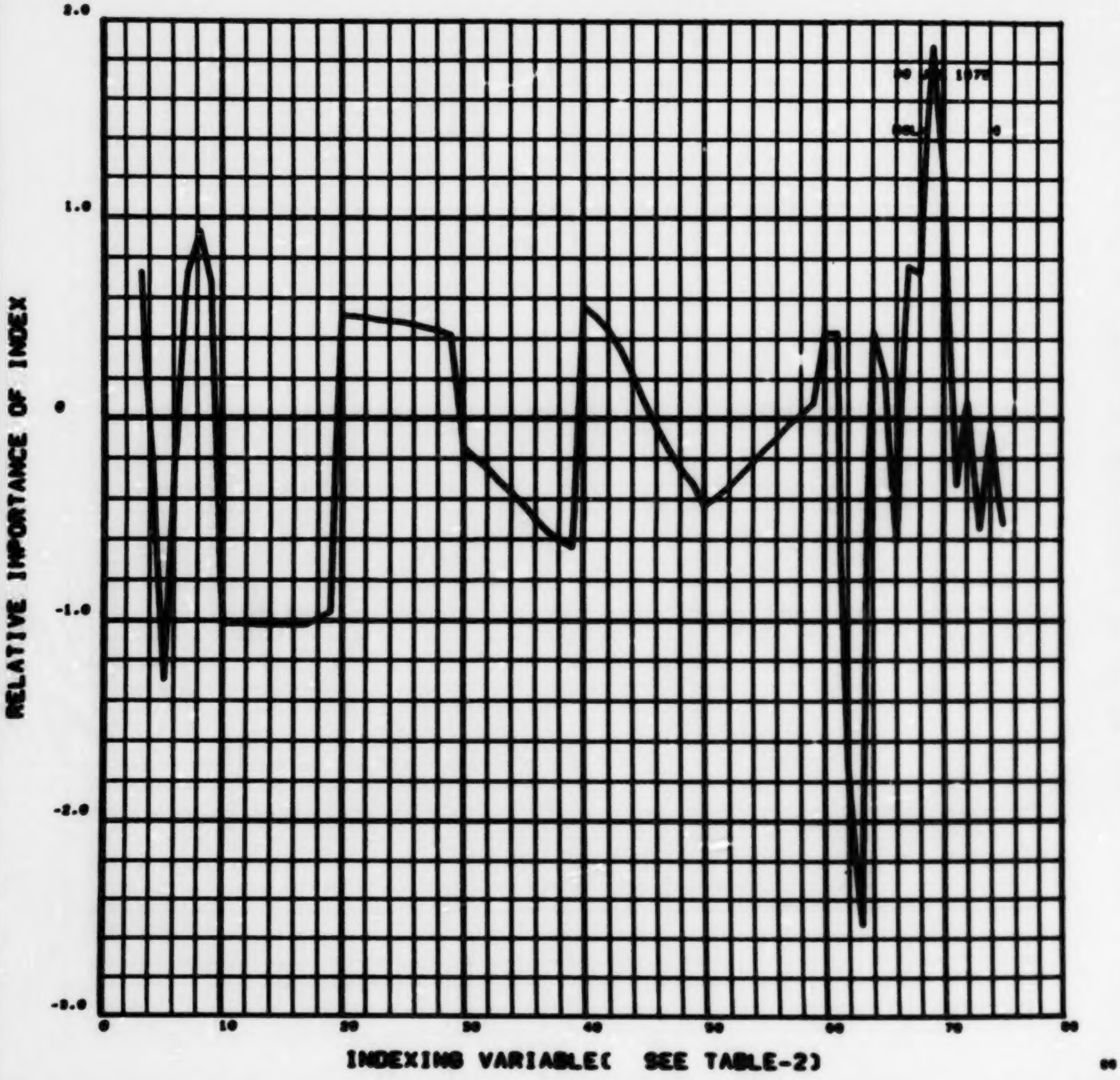


FIG- 5CRELATIVE IMPORTANCE VECTOR FOR AMBIENT CONC AT STATION-6 (N=16)

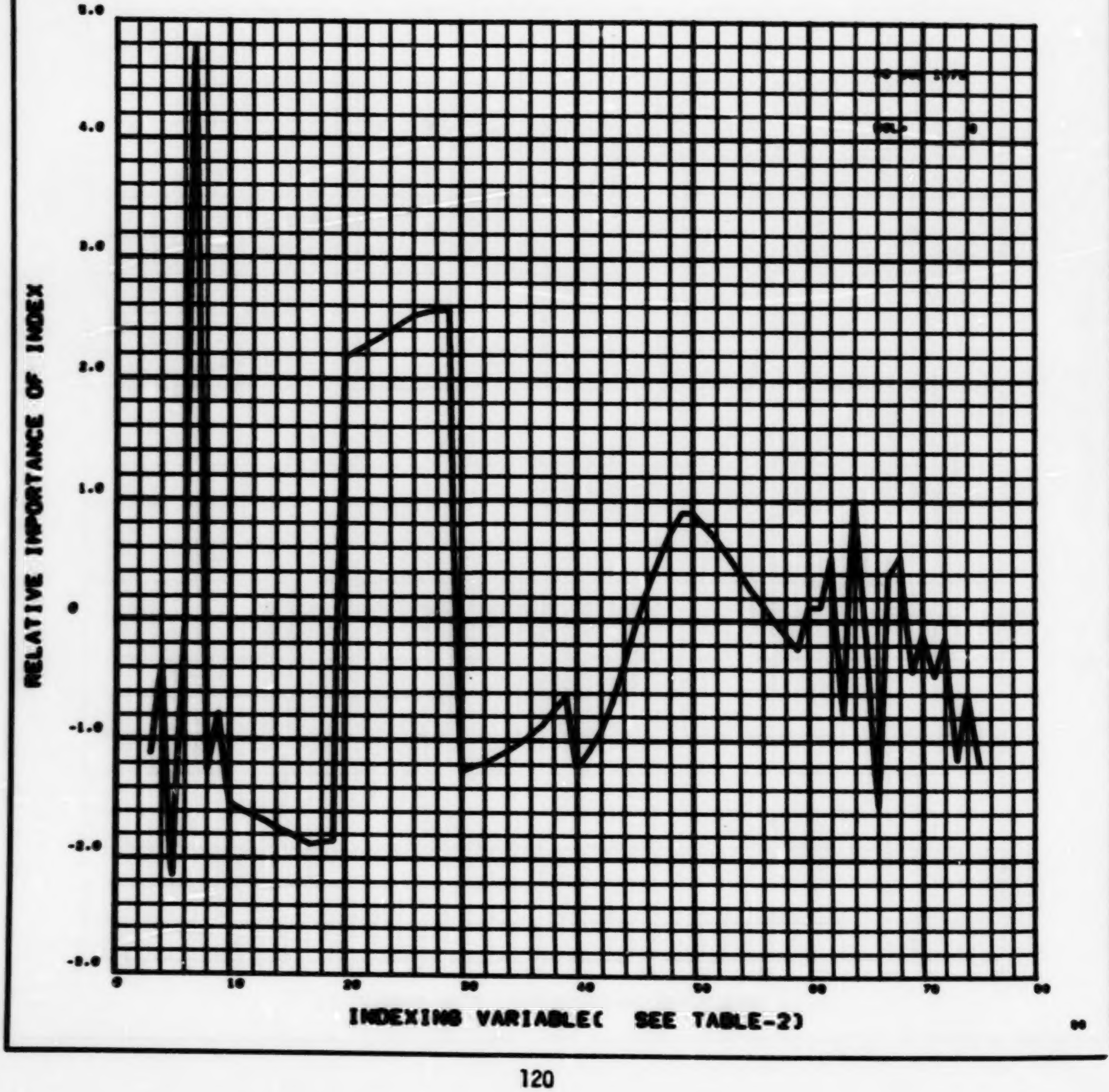


FIG- 6C RELATIVE IMPORTANCE VECTOR FOR AMBIENT CONC AT STATION-7 (N=12)

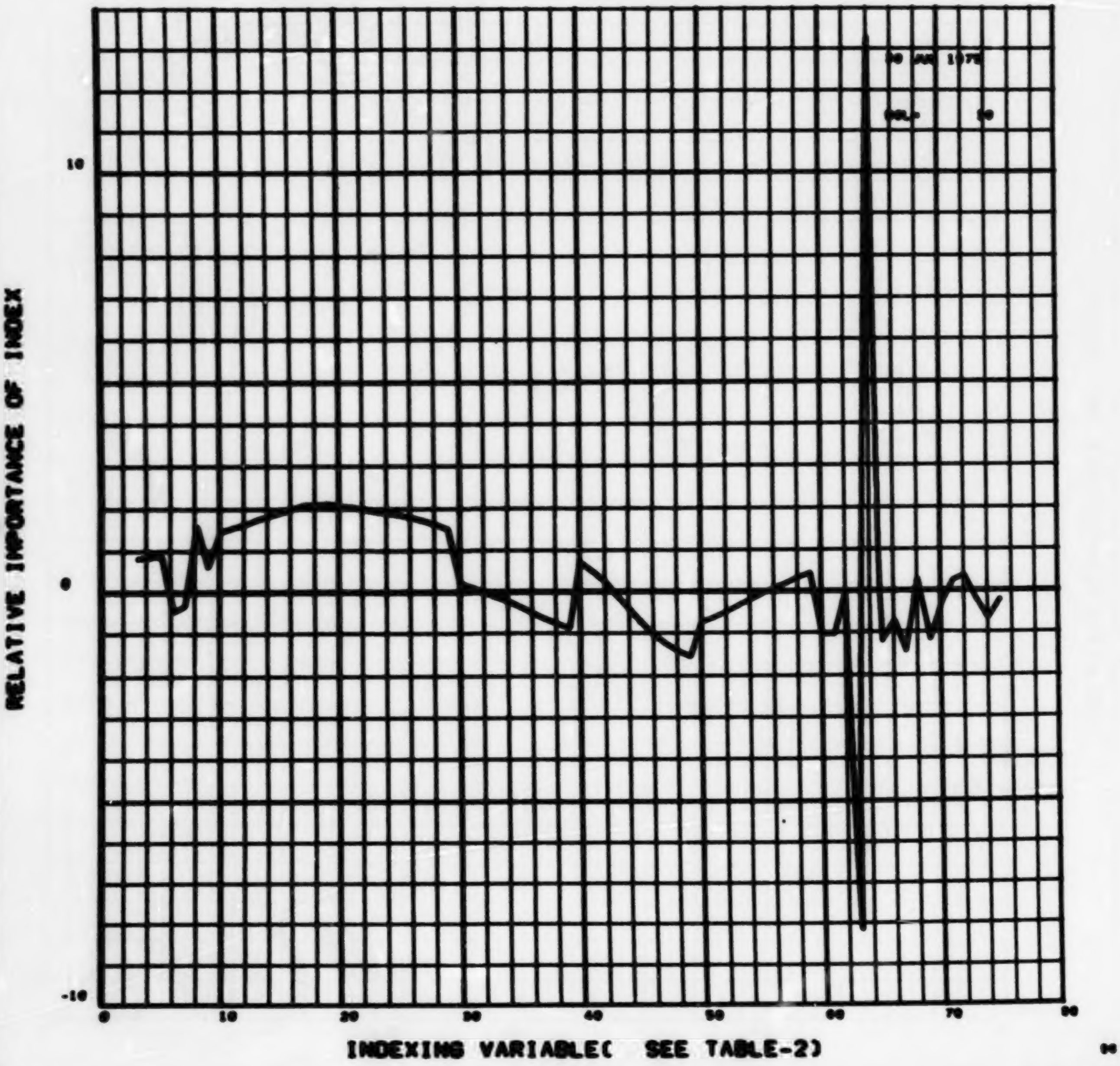
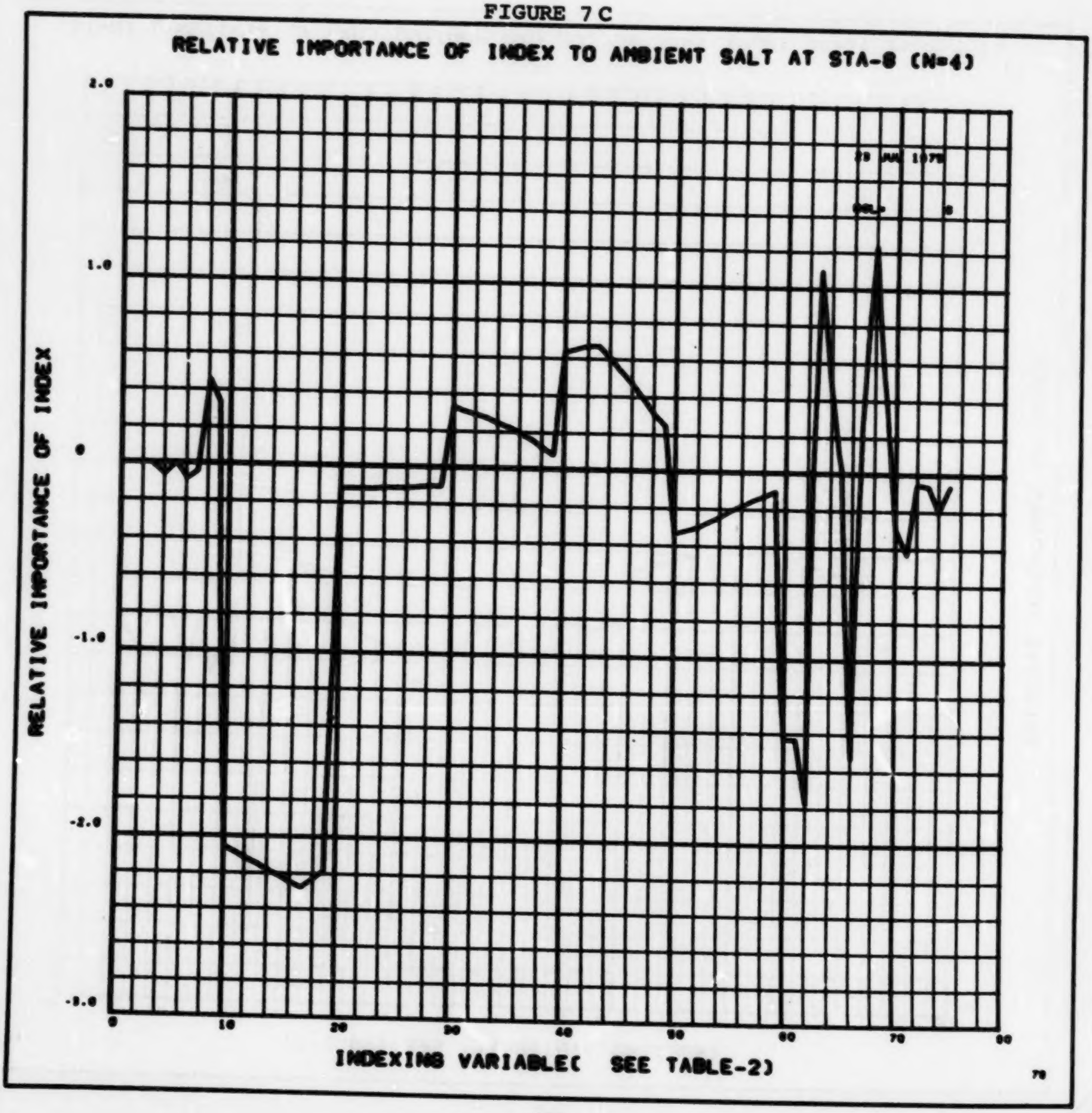


FIGURE 7C



RELATIVE IMPORTANCE OF INDEX TO AMBIENT SALT AT STA-8 (N=4)

FIGURE 8 C

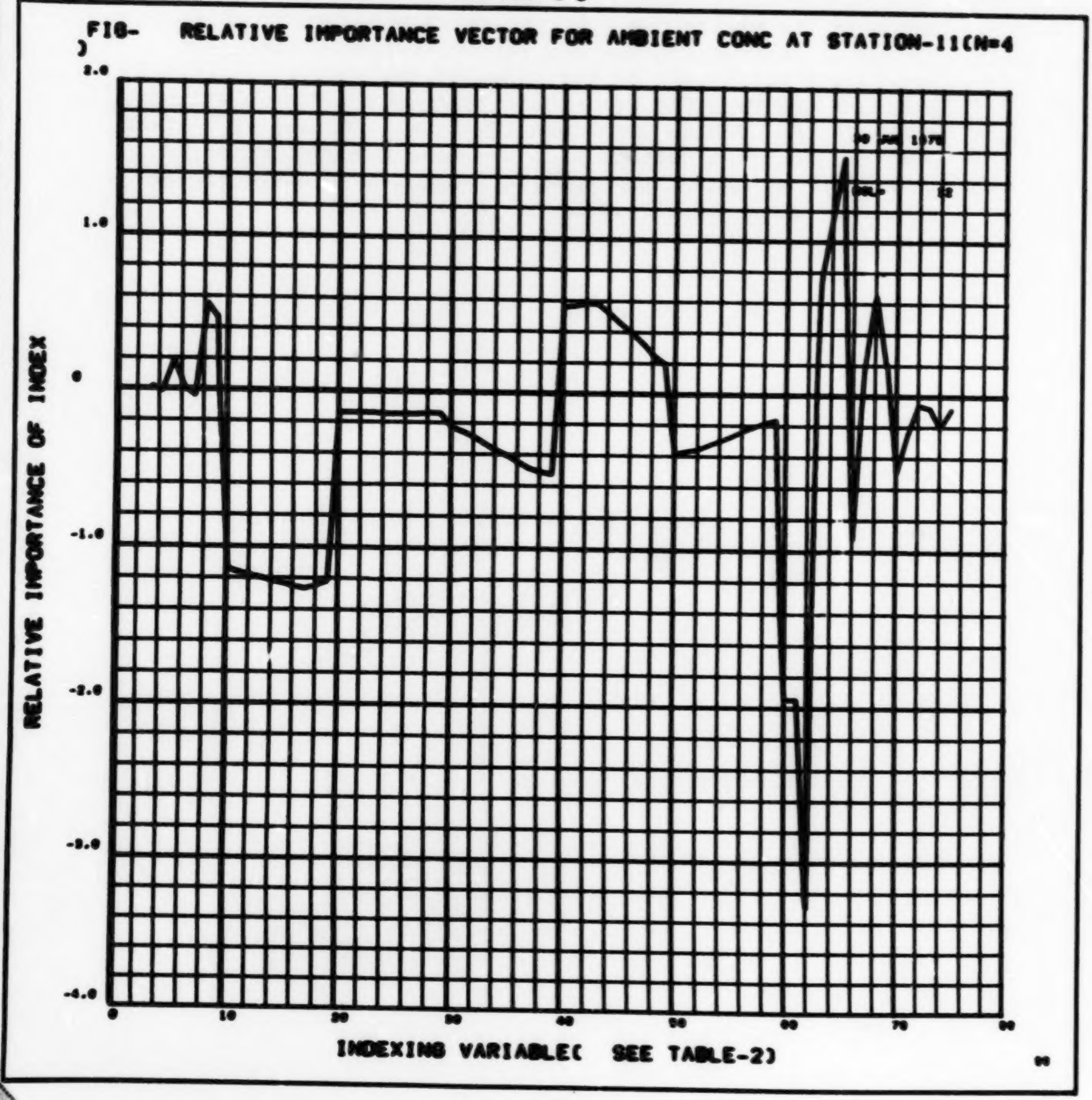


FIG-)
RELATIVE IMPORTANCE VECTOR FOR AMBIENT CONC AT STATION-11(N=4
2.0
1.0
0
-1.0
-2.0
-3.0
-4.0
RELATIVE IMPORTANCE OF INDEX
0
10
20
30
40
50
60
70
80
INDEXING VARIABLE(SEE TABLE-2)

APPENDIX D

ALGORTHIMS FOR CALCULATING AMBIENT CONCENTRATION

Tables 1D through 12D provide the information required to apply the algorithms derived to calculate the background concentrations at Stations 3 through 11 which were used to estimate the background concentrations under the conditions for which the cooling devices were operating. Table 1-D provides a definition of the indexing Variable i, Table 2-D presents the Variable $\bar{V}_i$ and Table 3D the Variables $VMAX_i$ and $VMIN_i$. Tables 4D through 11D present the average, $SCAMB_k$ and the algorithm vectors A_{ki}. To find the concentration at Station k, $CAMB_k$, the numbers presented in these tables should be combined according to the equation:

$$CAMB_k = SCAMB_k - \sum_{i=1}^{75} A_{ki} (V_i - \bar{V}_i)$$

where

$$V_i = 1 + ((VD_i - VMIN_i) / (VMAX_i - VMIN_i))$$

TABLE 1-D

DEFINITION OF DATA VECTOR = VD_i

VARIABLE NO i	SYMBOL	DESCRIPTION
1	d_e	Projection of position vector on East direction
2	d_N	Projection of position vector on North direction
3	CC1	Binary Code for Light Rain
4	CC3	Binary Code for Bugs on Sample
5	CC4	Binary Code for Dust Contamination
6	CC5	Binary Code for Combination of Comments
7	CC9	Binary Code for White Caps
8	ts	Start Time
9	te	End Time
10-19	dwi (i=1,10)	Projection of Wind Vector on Position Direction -10 Samples Between ts and te
20-29	Nwi (i=1,10)	Projection of Wind Vector on Normal to Position Direction -10 Samples Between ts and te
30-39	Ti (i=1,10)	Dry Bulb Temperature -10 Samples Between ts and te
40-49	Di (i=1,10)	Difference Between Dry and Wet Bulb Temperature -10 Samples Between ts and te
50-59	Hi (i=1,10)	Relative Humidity -10 Samples Between ts and te
-	CDC1	Binary Variable Indicating Cooling Tower Operation
-	CDC2	Binary Variable Indicating Spray Modules Operating
60	DY	Day of Year
61	DFT	Days Since First Test
62	S1	Binary Variable Indicating Spring
63	S2	Binary Variable Indicating Summer
64	S3	Binary Variable Indicating Fall
65	M	Binary Variable-Test on Monday

TABLE 1-D

DEFINITION OF DATA VECTOR = VD_i (CONT'D)

VARIABLE NO i	SYMBOL	DESCRIPTION
66	T_u	Binary Variable-Test on Tuesday
67	W	Binary Variable-Test on Wednesday
68	Th	Binary Variable-Test on Thursday
69	F	Binary Variable-Test on Friday
70	S	Binary Variable-Test on Saturday
71	dw (-1)	Projection of Preceding Day's Average Wind Vector on Position Direction
72	dn (-1)	Projection of Preceding Day's Average Wind Vector on Normal to Position Direction
73	dwsp (-1)	Preceding Day's Spread in Wind Speed
74	dnsp (-1)	Preceding Day's Spread in Wind Direction
75	dw (-1)	Preceding Day's Standard Deviation of Wind Speed
-	du (-1)	Preceding Day's Standard Deviation of Wind Direction
-	PRE	Precision Run Error

TABLE- 2-D THE AVERAGE VECTOR

I	AIV(I)	I	AIV(I)	I	AIV(I)	I	AIV(I)	I	AIV(I)	I	AIV(I)
1	0.136D 01	2	0.167D 01	3	0.102D 01	4	0.101D 01	5	0.103D 01	6	0.100D 01
7	0.102D 01	8	0.125D 01	9	0.115D 01	10	0.150D 01	11	0.150D 01	12	0.149D 01
13	0.148D 01	14	0.147D 01	15	0.147D 01	16	0.146D 01	17	0.145D 01	18	0.146D 01
19	0.147D 01	20	0.155D 01	21	0.154D 01	22	0.154D 01	23	0.153D 01	24	0.153D 01
25	0.152D 01	26	0.151D 01	27	0.149D 01	28	0.148D 01	29	0.147D 01	30	0.176D 01
31	0.176D 01	32	0.176D 01	33	0.176D 01	34	0.175D 01	35	0.174D 01	36	0.173D 01
37	0.171D 01	38	0.169D 01	39	0.167D 01	40	0.137D 01	41	0.138D 01	42	0.138D 01
43	0.140D 01	44	0.143D 01	45	0.146D 01	46	0.149D 01	47	0.150D 01	48	0.149D 01
49	0.148D 01	50	0.175D 01	51	0.177D 01	52	0.178D 01	53	0.179D 01	54	0.177D 01
55	0.175D 01	56	0.174D 01	57	0.172D 01	58	0.171D 01	59	0.169D 01	60	0.161D 01
61	0.161D 01	62	0.156D 01	63	0.107D 01	64	0.120D 01	65	0.131D 01	66	0.120D 01
67	0.117D 01	68	0.119D 01	69	0.116D 01	70	0.109D 01	71	0.155D 01	72	0.152D 01
73	0.102D 01	74	0.174D 01	75	0.102D 01						

TABLE- 3-D THE VMAX(I) AND VMIN(I) VECTORS

I	VMIN(I)	VMAX(I)	I	VMIN(I)	VMAX(I)	I	VMIN(I)	VMAX(I)
1	-0.1350E 04	0.1655E 04	2	-0.2200E 04	0.7600E 03	3	0.0	0.1000E 01
4	0.0	0.1000E 01	5	0.0	0.1000E 01	6	0.0	0.1000E 01
7	0.0	0.1000E 01	8	0.7490E 03	0.2091E 04	9	0.1014E 04	0.3386E 04
10	-0.3316E 02	0.3156E 02	11	-0.3260E 02	0.3172E 02	12	-0.3204E 02	0.3199E 02
13	-0.3149E 02	0.3226E 02	14	-0.3093E 02	0.3253E 02	15	-0.3037E 02	0.3280E 02
16	-0.2981E 02	0.3307E 02	17	-0.2926E 02	0.3334E 02	18	-0.3093E 02	0.3361E 02
19	-0.3260E 02	0.3387E 02	20	-0.3750E 02	0.3440E 02	21	-0.3645E 02	0.3428E 02
22	-0.3541E 02	0.3417E 02	23	-0.3436E 02	0.3405E 02	24	-0.3332E 02	0.3393E 02
25	-0.3227E 02	0.3382E 02	26	-0.3122E 02	0.3390E 02	27	-0.3018E 02	0.3487E 02
28	-0.2917E 02	0.3583E 02	29	-0.2880E 02	0.3680E 02	30	0.8500E 01	0.3320E 02
31	0.9311E 01	0.3310E 02	32	0.1012E 02	0.3300E 02	33	0.1093E 02	0.3290E 02
34	0.1174E 02	0.3296E 02	35	0.1256E 02	0.3314E 02	36	0.1337E 02	0.3333E 02
37	0.1418E 02	0.3376E 02	38	0.1443E 02	0.3438E 02	39	0.1450E 02	0.3500E 02
40	0.6000E 00	0.9700E 01	41	0.9444E 00	0.9233E 01	42	0.1289E 01	0.8767E 01
43	0.1433E 01	0.8300E 01	44	0.1222E 01	0.8189E 01	45	0.9778E 00	0.8111E 01
46	0.7333E 00	0.8033E 01	47	0.4889E 00	0.8344E 01	48	0.2444E 00	0.8822E 01
49	0.9537E-06	0.9300E 01	50	0.0	0.9500E 02	51	0.0	0.9267E 02
52	0.0	0.9033E 02	53	0.0	0.8967E 02	54	0.0	0.9122E 02
55	0.0	0.9278E 02	56	0.0	0.9433E 02	57	0.0	0.9589E 02
58	0.0	0.9744E 02	59	0.0	0.9900E 02	60	0.2360E 03	0.5700E 03
61	0.1000E 01	0.3350E 03	62	0.0	0.1000E 01	63	0.0	0.1000E 01
64	0.0	0.1000E 01	65	0.0	0.1000E 01	66	0.0	0.1000E 01
67	0.0	0.1000E 01	68	0.0	0.1000E 01	69	0.0	0.1000E 01
70	0.0	0.1000E 01	71	-0.7419E 02	0.5889E 02	72	-0.7749E 02	0.7749E 02
73	0.3000E 01	0.1302E 04	74	-0.1200E 02	0.8000E 01	75	0.1400E 00	0.2968E 03

Reproduced from best available copy.

TABLE 4D- ALGORITHM FOR STATION 3 (16 DIMENSIONS)

$SCAMB_3$= 6.116851

i	A_{3i}	i	A_{3i}	i	A_{3i}
1	2.42546077D 00	26	-5.77298797D-01	51	9.04645694D-01
2	-4.96631265D 00	27	-5.64798259D-01	52	6.51484604D-01
3	-8.91349300D-01	28	-5.51727421D-01	53	3.77777736D-01
4	-1.20514681D-01	29	-5.33437956D-01	54	9.74823498D-02
5	-1.88719382D 00	30	-1.91855979D 00	55	-1.73329004D-01
6	1.13870194D-01	31	-1.80354907D 00	56	-4.35373613D-01
7	2.14620022D 00	32	-1.67927958D 00	57	-6.88753522D-01
8	-1.64860387D 00	33	-1.54477161D 00	58	-9.34172050D-01
9	-1.02905422D 00	34	-1.38837149D 00	59	-1.17179914D 00
10	1.94046728D 00	35	-1.21242789D 00	60	-9.77321486D-01
11	1.97301569D 00	36	-1.02558150D 00	61	-9.77321486D-01
12	2.00253674D 00	37	-8.16806911D-01	62	7.27458417D-01
13	2.03222009D 00	38	-5.76968049D-01	63	-3.44744687D-01
14	2.06227453D 00	39	-3.42591873D-01	64	3.06231744D 00
15	2.09261487D 00	40	-2.01067788D 00	65	-1.21209601D 00
16	2.12321546D 00	41	-1.67049351D 00	66	-1.58922840D 00
17	2.15371250D 00	42	-1.25636689D 00	67	-4.59490257D-01
18	2.10936124D 00	43	-7.19777257D-01	68	2.34905064D-01
19	2.06754866D 00	44	-7.03490223D-02	69	1.72403330D-01
20	-5.95381628D-01	45	5.55435941D-01	70	1.68760144D-01
21	-5.92855891D-01	46	1.15264176D 00	71	6.08663796D-01
22	-5.90369861D-01	47	1.63783347D 00	72	-3.52706584D-01
23	-5.87663753D-01	48	2.01902576D 00	73	-9.26049133D-01
24	-5.84928356D-01	49	2.34092212D 00	74	-6.20483045D-01
25	-5.82094967D-01	50	1.14532684D 00	75	-9.33087957D-01

TABLE 5D- (16 DIMENSIONS) STATION 4 $SCAMB_4$ = 5.795147

i	A_{4i}	i	A_{4i}	i	A_{4i}
1	-1.60476817D 01	26	8.57624871D-01	51	-9.87522842D-01
2	-3.23332704D 01	27	7.81133134D-01	52	-9.81964400D-01
3	1.95971507D 00	28	7.04714597D-01	53	-9.57846042D-01
4	4.93630355D-01	29	6.22064905D-01	54	-9.10812825D-01
5	-1.24511573D 00	30	-3.82037717D-01	55	-8.64857236D-01
6	-4.29982948D-01	31	-4.11101552D-01	56	-8.21257738D-01
7	9.12103617D-01	32	-4.41635792D-01	57	-7.78379694D-01
8	1.28712292D 00	33	-4.75527059D-01	58	-7.37427351D-01
9	1.14152726D 00	34	-5.07939864D-01	59	-6.97298049D-01
10	-1.77463994D 00	35	-5.39387105D-01	60	3.37054285D-01
11	-1.72659467D 00	36	-5.72969335D-01	61	3.37054285D-01
12	-1.67441184D 00	37	-6.01439153D-01	62	6.99808430D-01
13	-1.62267708D 00	38	-6.07372354D-01	63	-3.13200872D 00
14	-1.57006485D 00	39	-6.07313376D-01	64	4.75755515D-01
15	-1.51638385D 00	40	7.47489023D-01	65	-4.08747097D-01
16	-1.46239796D 00	41	7.70121110D-01	66	-4.98933727D-01
17	-1.40824997D 00	42	7.96950097D-01	67	1.41062431D 00
18	-1.30651331D 00	43	8.06923872D-01	68	1.80577197D-01
19	-1.21076462D 00	44	7.34971071D-01	69	1.34045971D 00
20	1.19548951D 00	45	6.58722667D-01	70	5.63372007D-01
21	1.14437545D 00	46	5.85958628D-01	71	-6.12978796D-01
22	1.09090530D 00	47	4.90863728D-01	72	2.26522530D-01
23	1.03624439D 00	48	4.00725718D-01	73	-1.76702843D 00
24	9.79563476D-01	49	3.24349484D-01	74	-6.43405277D-01
25	-9.20915132D-01	50	-9.93267344D-01	75	-1.72088997D 00

TABLE 6D- (16 DIMENSIONS) STATION 5 $SCAMB_5 = 4.91$

i	A_{5i}	i	A_{5i}	i	A_{5i}
1	-3.33253519D 00	26	4.68274980D-01	51	-3.88432601D-01
2	-5.57589773D 00	27	4.52176599D-01	52	-3.38237678D-01
3	7.27794518D-01	28	4.35976205D-01	53	-2.79995408D-01
4	-2.13340726D-01	29	4.15695418D-01	54	-2.15626567D-01
5	-1.29410470D 00	30	-1.49560762D-01	55	-1.53270588D-01
6	-1.16232360D-01	31	-1.99434004D-01	56	-9.31573656D-02
7	7.26561543D-01	32	-2.53127250D-01	57	-3.48509930D-02
8	9.29432755D-01	33	-3.11441268D-01	58	2.14950327D-02
9	6.78530171D-01	34	-3.71930193D-01	59	7.61324350D-02
10	-1.01744160D 00	35	-4.34111133D-01	60	4.24794499D-01
11	-1.02031512D 00	36	-5.00101582D-01	61	4.24794499D-01
12	-1.02137330D 00	37	-5.63608899D-01	62	-1.80910435D 00
13	-1.02258704D 00	38	-6.05852249D-01	63	-2.52903040D 00
14	-1.02374602D 00	39	-6.40566512D-01	64	4.27678632D-01
15	-1.02477156D 00	40	5.58863261D-01	65	2.11950534D-01
16	-1.02590749D 00	41	5.04836720D-01	66	-6.04377566D-01
17	-1.02698757D 00	42	4.38936386D-01	67	7.61794441D-01
18	-9.92543110D-01	43	3.46810676D-01	68	7.25334854D-01
19	-9.60119128D-01	44	2.12436133D-01	69	1.87256754D 00
20	5.13926464D-01	45	8.11294059D-02	70	1.22291540D 00
21	5.07199276D-01	46	-4.41813949D-02	71	-3.30142092D-01
22	5.00089917D-01	47	-1.55841553D-01	72	7.84579452D-02
23	4.92909484D-01	48	-2.47764723D-01	73	-5.46610260D-01
24	4.85425738D-01	49	-3.25468796D-01	74	-6.39370100D-02
25	4.77624051D-01	50	-4.36248263D-01	75	-5.25412729D-01

TABLE 7D- (16 DIMENSIONS) STATION 6 $SCAMB_6 = 5.459854$

i	A_{6i}	i	A_{6i}	i	A_{6i}
1	5.84867587D 00	26	2.52109108D 00	51	7.86308133D-01
2	1.10258081D 01	27	2.54547157D 00	52	6.73042046D-01
3	-1.11381079D 00	28	2.56841554D 00	53	5.43436961D-01
4	-3.92593345D-01	29	2.56585914D 00	54	4.01915084D-01
5	-2.15098054D 00	30	-1.26650800D 00	55	2.64981558D-01
6	5.38434517D-02	31	-1.22917070D 00	56	1.32787241D-01
7	4.78532422D 00	32	-1.18914638D 00	57	4.71439155D-03
8	-1.25748313D 00	33	-1.14552092D 00	58	-1.19106774D-01
9	-7.86697345D-01	34	-1.09034119D 00	59	-2.39197165D-01
10	-1.54099437D 00	35	-1.02430728D 00	60	1.09462175D-01
11	-1.59139234D 00	36	-9.54073617D-01	61	1.09462175D-01
12	-1.63981646D 00	37	-8.68853515D-01	62	5.29247191D-01
13	-1.68836047D 00	38	-7.50565182D-01	63	-7.75753581D-01
14	-1.73744624D 00	39	-6.30753769D-01	64	1.00473025D 00
15	-1.78724136D 00	40	-1.23161789D 00	65	-1.23890360D-01
16	-1.83742111D 00	41	-1.09025071D 00	66	-1.54417706D 00
17	-1.88753585D 00	42	-9.17967742D-01	67	3.75493291D-01
18	-1.87156680D 00	43	6.83542383D-01	68	5.34301955D-01
19	-1.85652779D 00	44	3.62028321D-01	69	-4.21641141D-01
20	2.16829203D 00	45	4.90877328D-02	70	-1.09203325D-01
21	2.22338799D 00	46	2.49520990D-01	71	-4.55483189D-01
22	2.28052927D 00	47	5.08339538D-01	72	-1.39672773D-01
23	2.33941501D 00	48	7.18621619D-01	73	-1.14975028D 00
24	2.40037857D 00	49	8.96292248D-01	74	-6.06488746D-01
25	2.46343749D 00	50	8.94165961D-01	75	-1.18280227D 00

TABLE 8D - (12 DIMENSIONS) STATION 7 SCAMB$_7$ = 5.046803

i	A_{7i}	i	A_{7i}	i	A_{7i}
1	-5.89134390D 00	26	1.76010596D 00	51	-6.20216796D-01
2	-1.73470027D 01	27	1.67821369D 00	52	-4.98271360D-01
3	8.00221059D-01	28	1.59549484D 00	53	-3.62920435D-01
4	8.15547703D-01	29	1.49790401D 00	54	-2.20160465D-01
5	8.66471276D-01	30	1.85662322D-01	55	-8.19265882D-02
6	-4.63062906D-01	31	8.77855321D-02	56	5.13809277D-02
7	-3.39706609D-01	32	-1.74769244D-02	57	1.80642583D-01
8	1.55761793D 00	33	-1.31894589D-01	58	3.05524303D-01
9	5.96955537D-01	34	-2.53957641D-01	59	4.26722377D-01
10	1.43910405D 00	35	-3.82676114D-01	60	-1.05232853D 00
11	1.52663335D 00	36	-5.19483617D-01	61	-1.05232853D 00
12	1.61314941D 00	37	-6.57171666D-01	62	-5.41566243D-02
13	1.69980359D 00	38	-7.70228501D-01	63	-8.15917483D 00
14	1.78741608D 00	39	-8.70961319D-01	64	1.32690723D 01
15	1.87620201D 00	40	6.90053937D-01	65	-1.22039150D 00
16	1.96565601D 00	41	4.77275153D-01	66	-7.52401933D-01
17	2.05533014D 00	42	2.17957174D-01	67	-1.46004724D 00
18	2.07219412D 00	43	-1.00998445D-01	68	2.46961767D-01
19	2.08608964D 00	44	-4.33222526D-01	69	-1.17463340D 00
20	2.04953593D 00	45	-7.49016054D-01	70	-3.61623714D-01
21	2.00615675D 00	46	-1.05035731D 00	71	2.09731682D-01
22	1.96104459D 00	47	-1.27204984D 00	72	3.32897380D-01
23	1.91467603D 00	48	-1.43562948D 00	73	-2.47548793D-01
24	1.86664090D 00	49	-1.57426548D 00	74	-6.79397119D-01
25	1.81654613D 00	50	-7.36397816D-01	75	-2.59394553D-01

TABLE 9D- (4 DIMENSIONS) STATION 8 SCAMB$_8$ = 5.663548

i	A_{8i}	i	A_{8i}	i	A_{8i}
1	-6.33869465D-01	26	-1.10297053D-01	51	-3.18229512D-01
2	2.22795162D-01	27	-1.06760573D-01	52	-2.99457511D-01
3	-1.73989683D-02	28	-1.03160708D-01	53	-2.74498761D-01
4	-6.35142927D-02	29	-9.85901810D-02	54	-2.43068936D-01
5	1.08164161D-03	30	3.26652656D-01	55	-2.12711587D-01
6	-8.11876957D-02	31	3.09109185D-01	56	-1.83350606D-01
7	-4.40092867D-02	32	2.90171144D-01	57	-1.54942133D-01
8	4.48024787D-01	33	2.69669550D-01	58	-1.27444530D-01
9	3.27055201D-01	34	2.45570612D-01	59	-1.00802102D-01
10	-2.06188924D 00	35	2.18274675D-01	60	-1.43356166D 00
11	-2.09586349D 00	36	1.89260971D-01	61	-1.43356166D 00
12	-2.12655912D 00	37	1.56509919D-01	62	-1.77791361D 00
13	-2.15753494D 00	38	1.17784106D-01	63	1.09210798D 00
14	-2.18879162D 00	39	7.97079943D-02	64	4.43258288D-01
15	-2.22034656D 00	40	6.33366645D-01	65	2.21988650D-02
16	-2.25217795D 00	41	6.48336527D-01	66	-1.52869957D 00
17	-2.28393597D 00	42	6.66560216D-01	67	3.63871226D-01
18	-2.23626953D 00	43	6.69126835D-01	68	1.23338082D 00
19	-2.19138795D 00	44	6.03594940D-01	69	3.92427912D-01
20	-1.19809473D-01	45	5.34860991D-01	70	-2.99710691D-01
21	-1.18404514D-01	46	4.69256711D-01	71	-4.34781780D-01
22	-1.16948390D-01	47	3.86468290D-01	72	-5.29709879D-02
23	-1.15449349D-01	48	3.08510079D-01	73	-6.74931844D-02
24	-1.13900934D-01	49	2.42649065D-01	74	-1.99453779D-01
25	-1.12273125D-01	50	-3.36068249D-01	75	-6.48748376D-02

TABLE 10D- (8 DIMENSIONS) STATION 9 $SCAMB_9 = 4.139158$

i	A_{9i}	i	A_{9i}	i	A_{9i}
1	-4.49663659D-02	26	2.31879252D 00	51	1.54295250D-01
2	-7.46958314D-01	27	2.31304965D 00	52	1.72156067D-01
3	1.07058601D-01	28	2.30608587D 00	53	1.87435827D-01
4	-9.46394474D-02	29	2.27655705D 00	54	1.97961415D-01
5	1.60746174D-01	30	1.74752327D-01	55	2.08173845D-01
6	-1.55399926D-01	31	1.64790652D-01	56	2.18015766D-01
7	-7.33086190D-02	32	1.54048761D-01	57	2.27553342D-01
8	2.77614409D-01	33	1.42391095D-01	58	2.36782529D-01
9	1.69036120D-01	34	1.28790869D-01	59	2.45721957D-01
10	-5.54367004D-01	35	1.13429258D-01	60	-1.20359985D 00
11	-5.52167867D-01	36	9.71534839D-02	61	-1.20359985D 00
12	-5.48999131D-01	37	7.88473730D-02	62	-1.88616703D 00
13	-5.45808707D-01	38	5.75322634D-02	63	3.56292429D-01
14	-5.42584407D-01	39	3.66664705D-02	64	1.13597702D 00
15	-5.39295958D-01	40	-1.35899224D-01	65	3.38347162D-01
16	-5.36017888D-01	41	-1.65883265D-01	66	4.18734861D-01
17	-5.32627376D-01	42	-2.02402164D-01	67	1.09052718D 00
18	-5.10962246D-01	43	-2.40545059D-01	68	3.41334281D-01
19	-4.90557527D-01	44	-2.56952335D-01	69	-1.26385525D 00
20	2.14664076D 00	45	-2.70326098D-01	70	-2.67422078D-01
21	2.17406679D 00	46	-2.83091226D-01	71	-2.81401830D-01
22	2.20238538D 00	47	-2.80684231D-01	72	3.01810801D-01
23	2.23169178D 00	48	-2.73180360D-01	73	-4.46681181D-02
24	2.26200045D 00	49	-2.66843374D-01	74	-1.97659178D-01
25	2.29334846D 00	50	1.37297689D-01	75	-4.77784267D-02

TABLE 11D- (12 DIMENSIONS) STATION 10 $SCAMB_{10} = 3.918661$

i	A_{10i}	i	A_{10i}	i	A_{10i}
1	1.54383943D 00	26	1.75529287D 00	51	2.72853496D-01
2	1.52115087D 00	27	1.76625763D 00	52	2.38392091D-01
3	8.14643000D-02	28	1.77632629D 00	53	1.98355233D-01
4	2.51908622D-01	29	1.76864126D 00	54	1.53857248D-01
5	3.50393505D-02	30	-7.51281362D-01	55	1.10848052D-01
6	1.48539867D-01	31	-7.33918444D-01	56	6.92734770D-02
7	1.58156644D-01	32	-7.15215815D-01	57	2.90354673D-02
8	-4.59069261D-01	33	-6.94914565D-01	58	-9.89726951D-03
9	-1.93299746D-01	34	-6.67974970D-01	59	-4.76408852D-02
10	2.57443117D 00	35	-6.34864400D-01	60	1.12383262D 00
11	2.61954819D 00	36	-5.99658279D-01	61	1.12383262D 00
12	2.66057311D 00	37	-5.55563142D-01	62	3.96737925D-01
13	2.70200859D 00	38	-4.90327668D-01	63	1.43268452D 00
14	2.74381828D 00	39	-4.23524018D-01	64	-4.54523075D 00
15	2.78597582D 00	40	-6.13516737D-01	65	8.65211356D-01
16	2.82852307D 00	41	-5.89717749D-01	66	3.98986511D-02
17	2.87102705D 00	42	-5.60724384D-01	67	4.18711725D-01
18	2.81360824D 00	43	-5.09413906D-01	68	1.12627721D 00
19	2.75952644D 00	44	-4.02323097D-01	69	1.41250984D-01
20	1.54186402D 00	45	-2.95475207D-01	70	1.18721804D-02
21	1.57533290D 00	46	-1.93497962D-01	71	4.63828836D-01
22	1.60992119D 00	47	-9.13303768D-02	72	3.21505180D-01
23	1.64568517D 00	48	-2.62524642D-03	73	5.43391289D-02
24	1.68268622D 00	49	7.23110266D-02	74	2.77226981D-02
25	1.72095166D 00	50	3.05643910D-01	75	6.19849770D-02

TABLE 12D- (4 DIMENSIONS) STATION 11 $SCAMB_{11}$ = 7.388696

i	A_{11i}	i	A_{11i}	i	A_{11i}
1	-5.86323127D-02	26	-1.40764309D-01	51	-3.70228406D-01
2	7.50977410D-02	27	-1.39522641D-01	52	-3.54267707D-01
3	1.57492110D-02	28	-1.38157161D-01	53	-3.31168261D-01
4	-1.35168326D-02	29	-1.35449666D-01	54	-3.00244462D-01
5	1.88676914D-01	30	-2.21933040D-01	55	-2.70340534D-01
6	2.26379453D-03	31	-2.54711443D-01	56	-2.41451323D-01
7	-3.73047853D-02	32	-2.90091914D-01	57	-2.13477296D-01
8	5.51053279D-01	33	-3.28407580D-01	58	-1.86413367D-01
9	4.65981364D-01	34	-3.67335695D-01	59	-1.60189992D-01
10	-1.15121175D 00	35	-4.06478829D-01	60	-1.95206617D 00
11	-1.17022082D 00	36	-4.48057656D-01	61	-1.95206617D 00
12	-1.18740365D 00	37	-4.86438852D-01	62	-3.29588364D 00
13	-1.20474748D 00	38	-5.06459910D-01	63	7.51950727D-01
14	-1.22224206D 00	39	-5.20909973D-01	64	1.17942577D 00
15	-1.23989096D 00	40	5.59659353D-01	65	1.54421644D 00
16	-1.25771539D 00	41	5.70344776D-01	66	-9.20644355D-01
17	-1.27550368D 00	42	5.83337482D-01	67	1.63442863D-01
18	-1.24890902D 00	43	5.82045362D-01	68	6.45855191D-01
19	-1.22387818D 00	44	5.21249949D-01	69	2.03784184D-01
20	-1.35383236D-01	45	4.57852559D-01	70	-4.91045365D-01
21	-1.36239906D-01	46	3.97347772D-01	71	-2.36982827D-01
22	-1.37181792D-01	47	3.22721877D-01	72	-6.13333172D-02
23	-1.38124253D-01	48	2.52569204D-01	73	-8.27530747D-02
24	-1.39123788D-01	49	1.94023043D-01	74	-2.02868086D-01
25	-1.40147733D-01	50	-3.85413706D-01	75	-8.38596629D-02